Analytical Tools for Real-Time Delta-Sigma Multibeam Processing

Final Technical Report

December 2004

Ronald Coifman and Vladimir Rokhlin

of the

Yale University Department of Mathematics

Principal Investigator:

Vladimir Rokhlin
(203) 432-1278
rokhlin@cs.yale.edu

Sponsored by

Office of Naval Research
Contract # N00014-01-1-0364

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
20/12/2004	final	02-FEB-2001 to 01-JUN-2004

4. TITLE AND SUBTITLE
Analytical Tools for Real-Time Delta-Sigma Multibeam Processing

5a. CONTRACT NUMBER
N00014-01-1-0364

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Vladimir Rokhlin

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Yale University / Grant & Contract Administration
POB 208337 / 155 Whitney Avenue, Room 214
New Haven CT 06520-8337

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-1196, RR-1251

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 North Quincy Street
Arlington VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; distribution is Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We apply the theory of band-limited functions and related techniques to two aspects of Real-Time Delta-Sigma Multibeam Processing: the design of complicated antenna patterns and the improvement of delta-sigma algorithms in environments involving multiple transducers. We also apply modern fast electromagnetic modeling techniques to the simulation of antenna arrays, which is a critical step in the design of large-scale tightly packed structures.

15. SUBJECT TERMS
band-limited signals, antenna arrays, beam forming, compression of low-rank matrices, skeletonization. optimal antenna patterns

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Vladimir Rokhlin
			UU	126	19b. TELEPONE NUMBER *(Include area code)* (203) 432-1278

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

Summary

In constructing the suite of tools for the real-time delta-sigma multibeam processing, we performed the following steps.

1. The algorithm for the design of antenna arrays (corresponding to user-prescribed antenna patterns) has been improved. While certain generalizations of classical results (Markov, Karlin, et alii) state that certain antenna structures exist, algorithms for constructing such structures often have a heuristic element. Our older schemes have a non-trivial failure rate. The rate has been sufficiently low that practically any feasible structure can be designed, and sufficiently high to require human intervention and make the design a potentially time-consuming process.

The new algorithm has a negligible failure rate, and the very definition of failure is much more benign: instead of failing to produce the required structure, the scheme (very rarely) returns a structure containing one extra element (compared to the optimal one).

2. When a single antenna array has to perform several tasks, the problem is (at the present time) approached via simple scheduling: the tasks are ordered in terms of their urgency, and performed one after another. Obviously, some of the tasks can be performed simultaneously, either with the existing hardware or with very minor modifications. Other tasks might be fundamentally incompatible, and have to be scheduled no matter what equipment is to be used. We have developed criteria (applied to far-field patterns) that permit us to avoid attempts to design impossible structures. There appears to be a chance that such techniques could be used for real-time scheduling.

3. We discovered that the CPU time requirements associated with the design of the required patterns for real-life antenna arrays (assuming the use of classical algorithms from linear algebra) exceeds the capabilities of the available computer hardware; Moore's law alone is unlikely to alter this situation in the near future. This had been expected, and the effort was initiated to design and build the requisite analytical and numerical tools.

4. We discovered a new class of algorithms for the "compression" of low-rank matrices; we refer to the approach as "skeletonization". It is somewhat similar to the classical singular value decomposition (SVD) but requires considerably less CPU time to construct, and leads to schemes for the direct (as opposed to iterative) solution of systems of linear algebraic equations as-

sociated with the numerical scattering theory, design of antenna arrays and patterns, etc.

5. It has been observed that the new approach might lead to new "fast" algorithms for the diagonalization of (and for the construction of the SVD for) certain classes of matrices, including those encountered in the design of optimal antenna patterns for conformal and volume antenna arrays. Rudimentary codes of this type have been constructed.

Whenever physical signals are measured or generated, the locations of receivers or transducers have to be selected. Most of the time, this appears to be done on an ad hoc basis. For example, when a string of geophones is used in the measurements of seismic data in oil exploration, the receivers are located at equispaced points on an interval. When phased array antennae are constructed, their shapes are determined by certain aperture considerations; round and rectangular shapes are common. When antenna beams are steered electronically, it is done by changing the phases (and sometimes, the amplitudes) of the transducers. Again, these transducers are located in a region of predetermined geometry, and their actual locations within that geometry are chosen via some heuristic procedure.
In all these (and many other) cases, the signals being received or generated are *band-limited.* Optimal representation of such signals has been studied in detail by Slepian et. al. more than 30 years ago, and some of the obtained results were applied by D. Rhodes to the design of antenna patterns; further development of this line of research appears to have been hindered by the absence at the time of necessary numerical tools. We combine these classical results with the recently developed apparatus of Generalized Gaussian Quadratures to construct optimal nodes for the measurement and generation of band-limited signals. In this report, we describe the procedure based on these techniques for the design of such receiver (and transducer) configurations in a variety of environments.

A Procedure for the Design of Apparata for the Measurement and Generation of Band-Limited Signals

V. Rokhlin
Research Report YALEU/DCS/RR-1196
March 29, 2000

The author was supported in part by DARPA/AFOSR under Contract F49620/97/1/0011, in part by ONR under Grant N00014-96-1-0188, an in part by AFOSR under STTR number F49620/98/C/0051
Approved for public release: distribution is unlimited
Keywords: *Band-limited Signals, Antenna Arrays, Beam-forming*

1 Introduction

When measurements are performed, it often happens that the signal to be measured is well approximated by linear combinations of oscillatory exponentials, i.e. functions of the form

$$\sum_{j=1}^{n} \alpha_j \cdot e^{i \cdot \lambda_j \cdot x} \tag{1}$$

in one dimension, of the form

$$\sum_{j=1}^{n} \alpha_j \cdot e^{i \cdot (\lambda_j \cdot x + \mu_j \cdot y)} \tag{2}$$

in two dimensions, and of the form

$$\sum_{j=1}^{n} \alpha_j \cdot e^{i \cdot (\lambda_j \cdot x + \mu_j \cdot y + \nu_j \cdot z)} \tag{3}$$

in three dimensions. In most cases, the signal is band-limited, i.e. there exist such real positive a that all $1 \leq j \leq n$,

$$| \lambda_j | \leq a \tag{4}$$

in one dimension,

$$\lambda_j^2 + \mu_j^2 \leq a^2 \tag{5}$$

in two dimensions, and

$$\lambda_j^2 + \mu_j^2 + \nu^2 \leq a^2, \tag{6}$$

in three dimensions.

As is well-known, most measurements of electromagnetic and acoustic data (especially at reasonably high frequencies) are of this form. Examples of such situations include geophone and hydrophone strings in geophysics, phased array antennae in radar

systems, multiple transceivers in ultrasound imaging, and a number of other applications in astrophysics, medical imaging, non-destructive testing, etc.

In this report, we describe a procedure for determining the optimal distribution of sources and receivers that maximizes accuracy and resolution in measuring band-limited data given a fixed number of receivers. Alternatively, the procedure can be used to determine the optimal distribution of receivers that will minimize their number given specified accuracy and resolution. While the techniques described in this note are fairly general, we describe them in detail in the case of linear antenna arrays; the changes needed to generalize the approach to other cases are summarized in Section 6.

Remark 1.1 One of principal issues in the design of antenna arrays is the treatment (or avoidance) of the so-called supergain (or superdirectivity). Supergain is the condition that occurs when an antenna design is attempted that is prohibited (or nearly prohibited) by the Heisenberg principle; technically, it occurs in the form of very closely spaced elements operating out of phaze, and leads to prohibitive Ohmic losses in transmitting antennae, loss of sensitivity in receiving ones, etc. Since the purpose of this note is to introduce techniques for selecting the locations of elements *for a prescribed antenna pattern*, we avoid the issue of choosing the antenna pattern altogether. Instead, we observe design optimal element distributions for several standard far-field patterns (see Section 5.1), and we observe that the scheme for choosing optimal distributions of elements is virtually independent of the patterns being approximated.

Technically, the approach taken here is to observe that designing an antenna array can be viewed as constructing a quadrature formula for the integration of certain special classes of functions. Using recently developed techniques for the construction of so-called Generalized Gaussian Quadratures, we obtain both nodes and weights that are optimal (in a very strong sense) for the required antenna pattern.

The structure of this note is as follows. In Section 2, we summarize some of the mathematical apparatus to be used: Chebychev Systems, Generalized Gaussian Quadratures,

etc. In Section 3, we recapitulate some of the standard antenna theory, primarily to introduce the necessary notation. In Section 4, element distributions given a specific antenna pattern. In Section 5, we illustrate our approach with several numerical examples, and Section 6 contains a discussion of the generality of the schemes presented.

2 Analytical Preliminaries

In this section, we summarize several known facts about classical Special functions. All of these facts can be found in the literature; detailed references are given in the text.

2.1 Chebyshev systems

Definition 2.1 *A sequence of functions $\phi_1, \ldots, \phi_n$ will be referred to as a* Chebyshev system *on the interval $[a, b]$ if each of them is continuous and the determinant*

$$\begin{vmatrix} \phi_1(x_1) & \cdots & \phi_1(x_n) \\ \vdots & & \vdots \\ \phi_n(x_1) & \cdots & \phi_n(x_n) \end{vmatrix} \tag{7}$$

is nonzero for any sequence of points $x_1, \ldots, x_n$ such that $a \leq x_1 < x_2 \ldots < x_n \leq b$.

An alternate definition of a Chebyshev system is that any linear combination of the functions with nonzero coefficients must have no more than n zeros.

Examples of Chebyshev and extended Chebyshev systems include the following (additional examples can be found in [8]).

Example 2.1 *The powers $1, x, x^2, \ldots, x^n$ form an extended Chebyshev system on the interval $(-\infty, \infty)$.*

Example 2.2 *The exponentials $e^{-\lambda_1 x}, e^{-\lambda_2 x}, \ldots, e^{-\lambda_n x}$ form an extended Chebyshev system for any $\lambda_1, \ldots, \lambda_n > 0$ on the interval $[0, \infty)$.*

Example 2.3 *The functions $1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots, \cos nx, \sin nx$ form a Chebyshev system on the interval $[0, 2\pi]$.*

Example 2.4 *Suppose that $c > 0$ is a real number, w is a positive function $[-1,1] \to \mathbb{R}$ such that $w \in c^1[-1,1]$ and $w(-x) = w(x)$ for all $x \in [-1,1]$, n is a natural number, and the operators $P, Q : L^2[-1,1] \to L^2[-1,1]$ are defined by the formulae*

$$P(\phi)(x) = \int_{-1}^{1} w(t) \cdot e^{i \cdot c \cdot x \cdot t} \cdot \phi(t)\ dt \tag{8}$$

$$Q = P^* \circ P. \tag{9}$$

Suppose further that $\phi_1, \phi_2, \ldots$ are the eigenfunctions of Q, $\lambda_1, \lambda_2, \ldots$ are the corresponding eigenvalues, and $\lambda_1 > \lambda_2 > \lambda_3 \ldots$. Then all eigenfunctions of Q (also known as the right singular vectors of P) can be chosen to be real. Furthermore, the functions $\phi_1, \phi_2, \ldots, \phi_n$ constitute a Chebychev system on the interval $[-1,1]$.

2.2 Generalized Gaussian quadratures

A quadrature rule is an expression of the form

$$\sum_{j=1}^{n} w_j \cdot \phi(x_j), \tag{10}$$

where the points $x_j \in \mathbb{R}$ and coefficients $w_j \in \mathbb{R}$ are referred to as the nodes and weights of the quadrature, respectively. They serve as approximations to integrals of the form

$$\int_a^b \phi(x) \cdot \omega(x) dx \tag{11}$$

with ω is an integrable non-negative function.

Quadratures are typically chosen so that the quadrature (10) is equal to the desired integral (11) for some set of functions, commonly polynomials of some fixed order. Of these, the classical Gaussian quadrature rules consist of n nodes and integrate polynomials of order $2n-1$ exactly. In [13], the notion of a Gaussian quadrature was generalized as follows:

Definition 2.2 *A quadrature formula will be referred to as* Gaussian *with respect to a set of $2n$ functions $\phi_1, \ldots, \phi_{2n} : [a,b] \to \mathbb{R}$ and a weight function $\omega : [a,b] \to \mathbb{R}^+$, if it consists of n weights and nodes, and integrates the functions ϕ_i exactly with the weight function ω for all $i = 1, \ldots, 2n$. The weights and nodes of a Gaussian quadrature will be referred to as Gaussian weights and nodes respectively.*

The following theorem appears to be due to Markov [15, 16]; proofs of it can also be found in [10] and [8] (in a somewhat different form).

Theorem 2.1 *Suppose that the functions* $\phi_1, \ldots, \phi_{2n} : [a, b] \to \mathbb{R}$ *form a Chebyshev system on* $[a, b]$. *Suppose in addition that* $\omega : [a, b] \to \mathbb{R}$ *is a non-negative integrable function* $[a, b] \to \mathbb{R}$. *Then there exists a unique Gaussian quadrature for the functions* $\phi_1, \ldots, \phi_{2n}$ *on* $[a, b]$ *with respect to the weight function* ω. *The weights of this quadrature are positive.*

Remark 2.1 While the existence of Generalized Gaussian Quadratures was observed more than 100 years ago, the constructions found in [15, 16], [3, 10], [7, 8] do not easily yield numerical algorithms for the design of such quadrature formulae; such algorithms have been constructed recently (see [13, 28, 2]). The version of the procedure found in [2] was used to produce the results presented in the Examples 5.1, 5.2, 5.3 in Section 5.1; the reader is referred to [2] for details.

Applying Theorem 2.1 to the Example 2.4, we obtain the following theorem.

Theorem 2.2 *Suppose that under the conditions of Example 2.4,* n *is even. Then there exist* $n/2$ *points* $t_1, t_2, \ldots, t_{n/2}$ *on the interval* $[-1, 1]$ *and positive real numbers* $w_1, w_2, \ldots, w_{n/2}$ *such that*

$$\int_{-1}^{1} w(t) \cdot \phi_i(t)\, dt = \sum_{j=1}^{n/2} w_j \cdot \phi_i(t_j), \tag{12}$$

for all $i = 1, 2, \ldots, n$, *with* $\phi_1, \phi_2, \ldots, \phi_n$ *the first* n *eigenfunctions of the operator* Q *defined in (9).*

Corollary 2.3 *The above theorem provides a tool for the efficient approximate evaluation of integrals of the form (12), as follows. Given a positive real* ϵ, *we construct the*

Singular Value Decomposition of the operator P defined in (8). Choosing n to be the smallest even integer such that

$$\sum_{j=n+1}^{\infty} \lambda_j^2 < \epsilon^2, \tag{13}$$

we construct an $n/2$-point quadrature that integrates n first right singular functions exactly (effective numerical schemes for the construction of such quadratures can be found in [13, 28, 2]). Now, we observe that due to the triangle inequality combined with the positivity of the obtained weights $w_1, w_2, \ldots, w_{n/2}$,

$$|\sum_{j=1}^{n/2} w_j \cdot e^{i \cdot c \cdot x \cdot t_j} - \int_{-1}^{1} w(x) \cdot e^{i \cdot c \cdot x \cdot t}\, dt| < \epsilon \tag{14}$$

for any $x \in [-1, 1]$.

Remark 2.2 The principal subject of this note is the fact that the pattern of an antenna array is formed by a physical process amounting to a hardware implementation of a quadrature formula for functions of the form (9). Thus, designing a configuration of elements for such an antenna is equivalent to constructing a quadrature formula for functions of the form(9), and can be achieved via the techniques described in [13, 28, 2]).

3 Elements of Antenna Theory

In this section, we summarize certain facts about the theory of linear antenna arrays; all of these facts are well-known, and can be found, for example, in [9].

3.1 Pattern of a linear array

A source distribution σ on the interval $[-1, 1]$ creates the far-field pattern $f : [0, \pi] \to \mathbb{C}$ given by the formula

$$f(\theta) = \int_{-1}^{1} \sigma(u) \cdot e^{i \cdot k \cdot u \cdot cos(\theta)}\, du, \tag{15}$$

where k is the free-space wavenumber, u is the point on the interval $[-1,1]$, and θ is the angle between the point on the horizon where the far field is being evaluated and the x-axis. It is customary to introduce the notation

$$x = cos(\theta), \tag{16}$$

and define the function $F : [-1,1] \to \mathbb{C}$ by the formula

$$F(x) = f(acos(x)). \tag{17}$$

Now, defining the operator $A : L^2[-1,1] \to L^2[-1,1]$ by the formula

$$A(\sigma)(x) = \int_{-1}^{1} \sigma(u) \cdot e^{i \cdot k \cdot u \cdot x} \, du, \tag{18}$$

we observe that

$$F = A(\sigma) = \int_{-1}^{1} \sigma(u) \cdot e^{i \cdot k \cdot u \cdot x} \, du. \tag{19}$$

The function F is usually more convenient to work with than f, and the following obvious lemma is the principal reason for this difference.

Lemma 3.1 *Suppose that $\sigma \in L^2[-1,1]$, the function $F \in L^2[-1,1]$ is defined by (19), α is a real number, and the function $\tilde{\sigma} \in L^2[-1,1]$ is defined by the formula*

$$\tilde{\sigma}(u) = e^{i \cdot \alpha \cdot u} \cdot \sigma(u). \tag{20}$$

Then

$$A(\tilde{\sigma})(x) = A(\sigma)(x - \alpha) \tag{21}$$

for all $x \in (-\infty, \infty)$. In other words, in order to translate the antenna pattern F (viewed as a function of $x = cos(\theta)$) by α, one has to multiply by $e^{i \cdot \alpha \cdot k}$ the source distribution σ generating the pattern F.

Observation 3.1 *While the obvious physical considerations lead to the antenna pattern F defined on the interval $[-1,1]$, the formulae (15), (17) also define naturally the extension of F to the function $\mathbb{R} \to \mathbb{C}$; in a mild abuse of notation, we will be denoting by F both the original mapping $[-1,1] \to \mathbb{C}$ and its extension to the mapping $\mathbb{R} \to \mathbb{C}$. Similarly, we will be denoting by A both the operator $L^2[-1,1] \to L^2[-1,1]$ defined by (18) and its natural extension mapping $L^2[-1,1] \to c^\infty(\mathbb{R})$. The restriction of F on $\mathbb{R}\backslash[-1,1]$ is referred to as the invisible spectrum of the source distribution σ and plays an important role in the antenna theory (this role is discussed briefly in the following subsection). By the same token, the restriction of F on the interval $[-1,1]$ is referred to as the visible spectrum.*

When an antenna array is implemented in hardware, it is (usually) constructed of a finite collection of elements, as opposed to being a continuous source distribution. Mathematically, it is equivalent to replacing the general function σ in (15), (19) with σ defined by the expression

$$\sigma(x) = \sum_{j=1}^{n} \beta_j \cdot \phi_j(u), \tag{22}$$

with $\phi_1, \phi_2, \ldots, \phi_n$ the source distributions generated by individual elements, and the coefficients $\beta_1, \beta_2, \ldots, \beta_n$ the intensities of the elements. As a rule, the elements are localized in space (i.e. the functions $\phi_1, \phi_2, \ldots, \phi_n$ are supported on small subintervals of $[-1,1]$), and very often, all of the elements are identical (i.e. the functions ϕ_j are translates of each other), so that

$$\phi_j(u) = \phi(u - u_j), \tag{23}$$

with ϕ the source distribution of a single element located at the point $u = 0$, and u_j the location of the element number j. Obviously, the far-field pattern of ϕ is given by the formula

$$F_\phi(x) = \int_{-1}^{1} \phi(u) \cdot e^{i \cdot k \cdot u \cdot x} \, du; \tag{24}$$

combining (24) with (22) and (23), we obtain the identity

$$\sigma(x) = \int_{-1}^{1} \phi(u) \cdot e^{i \cdot k \cdot u \cdot x} \, du \cdot \sum_{j=1}^{n} \beta_j \cdot e^{i \cdot k \cdot u_j \cdot x}, \tag{25}$$

known in the antenna theory as the principle of pattern multiplication.

Remark 3.2 The standard form of the principle of multiplication reads: "The field pattern of an array of nonisotropic but similar point sources is the product of the pattern of the individual source and the the pattern of an array of isotropic point sources, having the same locations, relative amplitudes and phases as the nonisotropic point sources" (see [9]). Needless to say, this is a special case of the well-known theorem from the theory of the Fourier Transform, stating that the Fourier transform of the product of two functions is the convolution of the Fourier Transforms of multiplicants.

4 Antenna Patterns and Corresponding Optimal Element Distributions

4.1 Characteristics of an antenna pattern

Depending on the situation, the design of an antenna array attempts to optimize certain characteristics of the resulting far-field pattern, subject to certain constraints on the number, power, etc. of the elements. Since the principal purpose of this note is to describe a technique for the selection of the *locations* of the elements that approximate a user-specified pattern, we could use any reasonable far-field pattern to be approximated. In subsection 4.2, 4.3, we construct optimal element distributions for the so-called sector patterns and cosecant pattern, respectively; a detailed discussion of these (and several other) pattern cans be found, for example in [14].

We will say that the antenna pattern has the ϵ-bandwidth b if

$$\int_{b \le \|x\| \le 1} |F(x)|^2 \, dx = \epsilon^2 \cdot \int_{-1}^{1} |F(x)|^2 \, dx \tag{26}$$

in other words, the proportion of the energy radiated outside the ϵ-beamwidth from the axis of the beam is equal to ϵ. The *supergain* of an antenna is defined (see, for example, [27]), as the ratio

$$\frac{\int_{-\infty}^{+\infty} |F(x)|^2 \, dx}{\int_{-1}^{1} |F(x)|^2 \, dx}. \tag{27}$$

The supergain (sometimes referred to as superdirectivity) measures the ratio of the energy associated with the total spectrum of the antenna to the energy in its visible spectrum; while detailed discussion of supergain and related issues is outside the scope of this note, we will observe that antenna arrays with large degrees of supergain would violate the uncertainty principle, and thus are physically impossible. Attempts to construct supergain antennae result in rapidly (exponentially) growing Ohmic losses, prohibitive accuracy requirements, extremely low bandwidth, etc. Thus, any potentially useful procedure for the design of antenna arrays has to limit the supergain of the resulting patterns.

4.2 Sector patterns

It is often desirable to construct antenna patterns that are as constant as possible within the main beam, and as small as possible outside it; in other words, ideally, the pattern would be defined by the formulae

$$F_b(x) = 1 \quad for \quad |x| \leq b, \tag{28}$$

$$F_b(x) = 0 \quad for \quad |x| > b, \tag{29}$$

with b a real number such that $0 < b \leq k$. Needless to say, the function F_b defined by the formulae (28), (29) is not band-limited, and some approximation has to be used. A standard procedure is to truncate the Fourier Transform of F_b, approximating it by the function $\tilde{F}_b$ defined by the formula

$$\tilde{F}_b(x) = \int_{-1}^{1} \frac{sin(b \cdot t)}{t} \cdot e^{i \cdot k \cdot x \cdot t} \tag{30}$$

(see, for example, [26]). An important special case occurs when $b = k$, with (30) assuming the form

$$\tilde{F}_k(x) = \int_{-1}^{1} \frac{sin(k \cdot t)}{t} \cdot e^{i \cdot k \cdot x \cdot t}; \tag{31}$$

obviously, the latter expression is a band-limited approximation of the δ-function. Another frequently encountered situation is that of $b = k/2$, so that (30) assumes the form

$$\tilde{F}_k(x) = \int_{-1}^{1} \frac{sin(\frac{k}{2} \cdot t)}{t} \cdot e^{i \cdot k \cdot x \cdot t}, \tag{32}$$

which is a band-limited approximation to the beam that is equal to 1 for $-1/2 < x < 1/2$ and to zero elsewhere.

In Section 4.4 below, we demonstrate optimal element configurations that produce approximations to the patterns (31), (32) with $k = 20\pi, 10\pi, 32.4676\pi$.

Remark 4.1 While (30) is by no means the only possible band-limited approximations to to F_b, it is quite satisfactory in most cases, in addition to being simple. Furthermore, the principal purpose of this note is to describe a technique for the selection of *locations* of the nodes, given a pattern to be approximated. Thus, we ignore the issue of the optimal choice of F_b.

4.3 Cosecant patterns

Another standard far-field radiation pattern is the so-called cosecant pattern (see, for example, [19]). Given two real numbers $0 < a < b < 1$, the cosecant pattern $F_{a,b}$ is defined by the formula

$$F_{a,b}(x) = \frac{1}{x} \tag{33}$$

for all $x \in [a, b]$, and

$$F_{a,b}(x) = 0 \tag{34}$$

for all $x \in ([-1,1] \setminus [a,b])$. Again, the function $F_{a,b}$ defined by the formulae (33), (34) is not band-limited, and can not be represented by the expression of the form (24). Before the scheme of this note can be applied to $F_{a,b}$, the latter has to be approximated with a band-limited function; as discussed in Section 4.1 above, if such an approximation is to be useful as an antenna pattern, its supergain factor has to be controlled. Fortunately, a procedure for such an approximation has been in existence for more than 35 years (see, [18]); the algorithm of [18] is a modification of the least-squares approach *permitting the user to limit the supergain factor of the obtained pattern explicitly.* At the time, the utility of the scheme of [18] was limited by the (perceived) difficulty in the numerical evaluation of Prolate Spheroidal Wave functions; given the present state of numerical analysis, this difficulty is non-existent, and it is this author's impression that the insights of [18], [19] deserve more attention than they have been receiving.

4.4 Optimal distributions of elements

In this subsection, we briefly describe an algorithm for the construction of optimal (in the sense defined below) element configurations for the generation of antenna patterns given by (15), of which the patterns (29)-(31) are special cases. As will be seen, the procedure is in fact applicable to the design of element configurations for very general far-field patterns.

We start with observing that (15) expresses the far-field pattern F as an integral over the interval $[-1,1]$ of functions of the form

$$\sigma(u) \cdot e^{i \cdot k \cdot x \cdot u}, \tag{35}$$

with $x = cos(\theta)$ determined by the direction θ in which the far-field is being evaluated. In other words, the problem of finding efficient antenna element distributions is equivalent to that of constructing quadrature formulae for integrals of the form (8), with

$$w(t) = \sigma(t). \tag{36}$$

In the cases when σ is non-negative everywhere on the interval $[-1, 1]$, Theorem 2.2 guarantees the existence of Generalized Gaussian Quadratures, and [13, 28]) provide a satisfactory numerical apparatus for the construction of such quadratures. Obviously, the patterns given by the formula (28) are not generated by non-negative source distributions, except when

$$b \leq \pi. \tag{37}$$

Thus, for these (and many other) patterns, the conditions of Theorem 2.2 are violated, and the existence of Generalized Gaussian Quadratures is not guaranteed. In our numerical experiments, the techniques of [2]) (after some tuning) have always been successful in finding the Gaussian quadratures for integrals of the form (28); some of our results are presented in Section 5 below.

5 Numerical Examples

In this section, we present examples of optimal element distributions generating the patterns of the preceding Section; all of the results presented here have been obtained numerically. Antenna patterns we present are compared to the antenna patterns given by uniform source distributions; configurations of elements approximating these antenna patterns are compared to equispaced distributions of elements generating the same antenna patterns.

5.1 Optimal distributions of elements

In this section, we demonstrate the results of the application of the techniques of Section 4.4 of this note to the types of antenna patterns described in the Sections 4.2, 4.3.

In all cases, we choose the size of an antenna array and a pattern to be reproduced, and use the scheme outlined in Section 4.4 to design a distribution of antenna elements (both the locations and the intensities) located within the chosen array that reproduces the required pattern. For comparison, we also generate optimal (in the least squares sense)

approximations to the desired pattern generated by equispaced elements located within the same array. Since the number of equispaced nodes required to obtain a reasonable approximation to the desired pattern is (in many cases) much greater than the number of optimally chosen nodes, for each example we demonstrate patterns generated by several such configurations. In this manner, the numbers of optimally chosen nodes necessary to obtain reasonable approximations to the desired patterns can be compared to the numbers of equispaced nodes required to obtain similar results.

5.1.1 Sector patterns

Example 5.1 *The first example we consider is of the pattern defined by the formula (32), with* $k = 62.8312$, *so that the size of the array is* 20 *wavelengths.*

In Figure 5, we display an approximation to the pattern obtained with 19 *elements, overlayed with the exact pattern; the locations of the elements are displayed in Figure 5a; the relative error of the obtained approximation is* 5.01%.

Similarly, in Figure 5g, we display the approximation to the pattern obtained with 21 *elements, overlayed with the exact pattern; the relative error of the obtained approximation is* 0.443%; *in Figure 5h, we display the the approximation obtained with* 17 *elements. In the latter case, the relative error of the obtained approximation is* 6.43%; *Figure 5i depicts the* 17*-node distribution producing the approximation illustrated in Figure 5h. Finally, Figure 5j contains a graph of the values of the sources located at the* 17 *nodes depicted in Figure 5i and generating the pattern shown in Figure 5h.*

For comparison, the optimal approximation obtained with 19, 24, 29, 31, *and* 34 *equispaced elements are displayed in Figures 5b, 5c, 5d, 5e, 5f, respectively; these are also overlayed with the exact pattern.*

Example 5.2 *Our second example is identical to the first one, with the exception that* $k = 31.416$, *so that the size of the array is* 10 *wavelengths.*

In Figure 6, we display an approximation to the pattern obtained with 9 *elements, overlayed with the exact pattern; the locations of the elements are displayed in Figure 6a; the relative error of the obtained approximation is* 11.2%.

Similarly, in Figure 6f, we display the approximation to the pattern obtained with 11 *elements, overlayed with the exact pattern; the relative error of the obtained approximation is* 0.600%.

For comparison, the optimal approximation obtained with 9, 14, 16, *and* 18 *equispaced elements are displayed in Figures 6b, 6c, 6d, 5e, respectively; these are also overlayed with the exact pattern.*

Example 5.3 *Our third example is identical to the preceding two, with the exception that* $k = 102$, *so that the size of the array is about* 32.45 *wavelengths.*

In Figure 7a, we display an approximation to the pattern obtained with 23 *optimally distributed elements, overlayed with the exact pattern and with the pattern obtained with* 23 *equispaced elements.*

The relative error of the obtained approximation is 5.4%; *needless to say, the error of the approximation obtained with the equispaced nodes is more than* 70%. *As can be seen from Figure 7c, the actual size of the obtained* 23*-element array is about* 21 *wavelengths; in other words, in order to obtain this precision, the array needs to be about* 2/3 *of the nominal (maximum permitted) length.*

In Figure 7b, we display the approximation to the pattern obtained with 42 *and* 48 *elements, overlayed with the exact pattern.*

It is worth noting that with 33 *optimally distributed elements, the pattern is approximated to the precision* 0.12%; *we do not display the obtained pattern since it is visually indistinguishable from the pattern being approximated.*

Example 5.4 *Our final example is somewhat different from the preceding ones, in that instead of approximating a sector pattern, we approximate a cosecant pattern (see (33), (34) in Subsection 4.3 above).*

In this example, we set

$$a = sin(15^\circ), \tag{38}$$

$$b = sin(75°), \tag{39}$$

and use the procedure of [18] to approximate $F_{a,b}$ with a band-limited function. The band-limit has been more or less arbitrarily set to 110, *resulting in an antenna array about* 35 *wavelengths in size, and the supergain factor of the approximation was set to* 1.1.

In Figure 8a, we display an approximation to the pattern obtained with 53 *optimally distributed elements, overlayed with the exact bandlimited pattern and with the pattern obtained with* 53 *equispaced elements.*

The relative error of the obtained approximation is 1.79%*; the error of the approximation obtained with the equispaced nodes is about* 42%.

In Figure 8b, we display the approximation to the pattern obtained with 47 *optimally distributed elements, overlayed with the exact pattern; the purpose of this final figure is to demonstrate the behavior of the scheme when the number of elements is insufficient (i.e. when the array is underresolved).*

It is worth noting that it takes about 70 *equispaced nodes to obtain the resolution obtained with* 47 *optimally chosen ones.*

The following observations can be made from Figures 5 - 8b, and from the more detailed numerical experiments performed by the author.

1. In order to obtain reasonable precision, the scheme requires about 1 point per wavelength in the antenna array; this is more or less independent from the structure of the beam *as long as the pattern is symmetric about the point* $x = 0$. This fact is observed numerically, even for modest numbers of nodes; for large-scale arrays, this statement (interpreted asymptotically) can be proved rigorously. For certain beam structures, the required number of nodes is even less (see Example 5.3). The reasons for these additional savings are subtle, and have to do with the fact that the continuous source distribution generating the pattern is relatively small on a large part of the antenna array; the algorithm of [2] takes advantage of this fact to reduce the number of nodes. When the beam is not symmetric about $x = 0$, the number of elements required does depend on

the structure of the pattern, and the dependence is fairly complicated. Generally, the improvement for non-symmetric beams is less than that for the symmetric ones.

2. The qualiative behavior of the scheme is similar to that of the Gaussian quadratures in that it displays no convergence at all until a certain minimum number of nodes is achieved; after that, the convergence is very fast. This behavior is not surprising, since the scheme *is* based on a Generalized Gaussian quadrature.

3. For the sector pattern with the sector $[-1/2, 1/2]$, the scheme reduces the required number of nodes by a factor of about 1.5 for small-scale problems, and roughly by a factor of 2 for large-scale ones; again, for large-scale problems, an asymptotic version of this statement can be proven rigorously.

4. For the cosecant pattern with the parameters specified by (38), (39), the number of nodes required is reduced by approximately a factor of 1.4. As the sidelobe level is reduced, the improvement obtained by going from the equispaced discretization to the optimal one increases rapidly.

5. An examination of Figures 5a, 6a shows that while the optimal nodes are by no means uniform, they display no clustering behavior.

6. An examination of Figure 5j shows that the intensities of individual elements do not become large; this is confirmed by the more extensive numerical experiments performed by the author.

7. The combination of the preceding two paragraphs (combined with additional numerical experiments and analysis) provide evidence that configurations of this type should pose no supergain problems.

6 Generalizations

The results described above admit radical generalizations in several directions; several such directions are discussed below,

1. Conformal one-dimensional arrays. The extension of the techniques of this note to one-dimensional arrays located on curves in R^3 is completely straightforward, involving only a modest increase of the CPU time requirements of the procedure. Improvement in the number of nodes required to produce a prescribed pattern is similar to that in the case of a linear array.

2. Planar two-dimensional arrays. A straightforward generalization of the results of Sections 4, 5, is to rectangular planar arrays. Here, a tensor product quadrature can be constructed from the quadratures of Sections 4, 5, possessing all of the desirable properties of the latter. Obviously, the advantage in the number of transducers is squared, so that (for example) replacing 50 nodes in each of the two directions by 23 nodes (see Example 5.3 above) will lead to a factor of $(50/23)^2 \sim 4.7$ savings in the number of elements.

The theory of Section 4 has been extended for disk-shaped arrays, via (*inter alia*) the techniques developed in [23]. The improvement in the number of nodes is comparable to that obtained in the rectangular geometry, and the CPU time requirements do not differ appreciably from those in the case of linear one-dimensional arrays.

The extension of the theory to more general geometries in the plane is in progress. At the present time, our only numerical experiments have been with arrays on triangles; the results are encouraging, but the CPU time requirements of the algorithms are excessive (we have only been able to design triangular arrays about 6 wavelengths in size). We are now in the process of constructing a more efficient numerical procedure for such computations.

3. Conformal two-dimensional arrays. The only environment in which we have a satisfactory theory is when the array is located on a surface of revolution; even in this environment, no experiments have been performed. We have not investigated more general conformal two-dimensional arrays in sufficient detail.

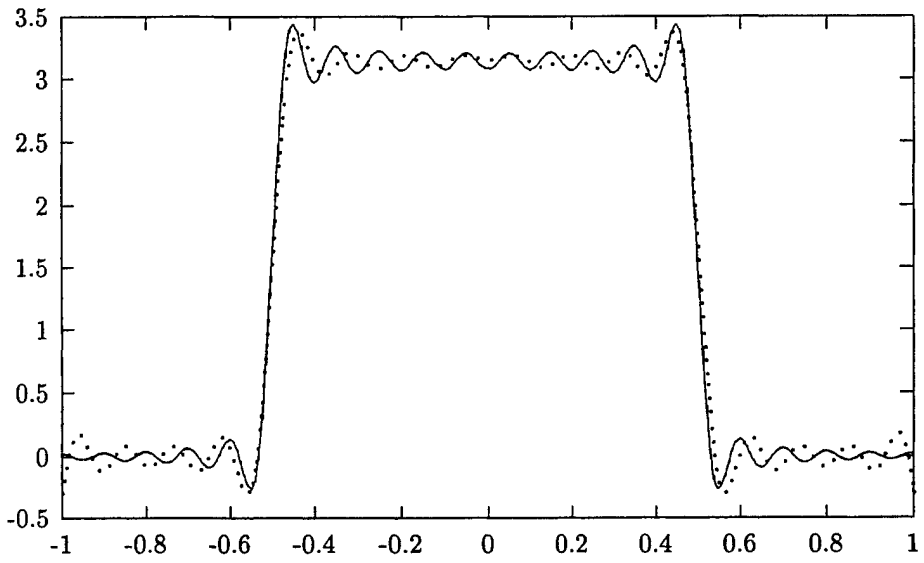


Figure 5: The pattern created by the 19 optimal elements, depicted in Figure 5a as described in Example 5.1

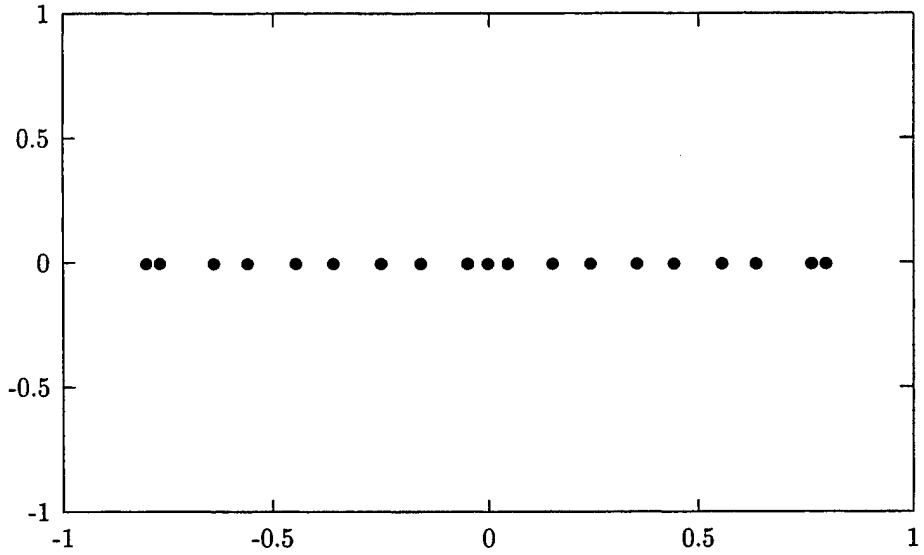


Figure 5a: The distribution of elements creating the pattern depicted in Figure 5, as described in Example 5.1

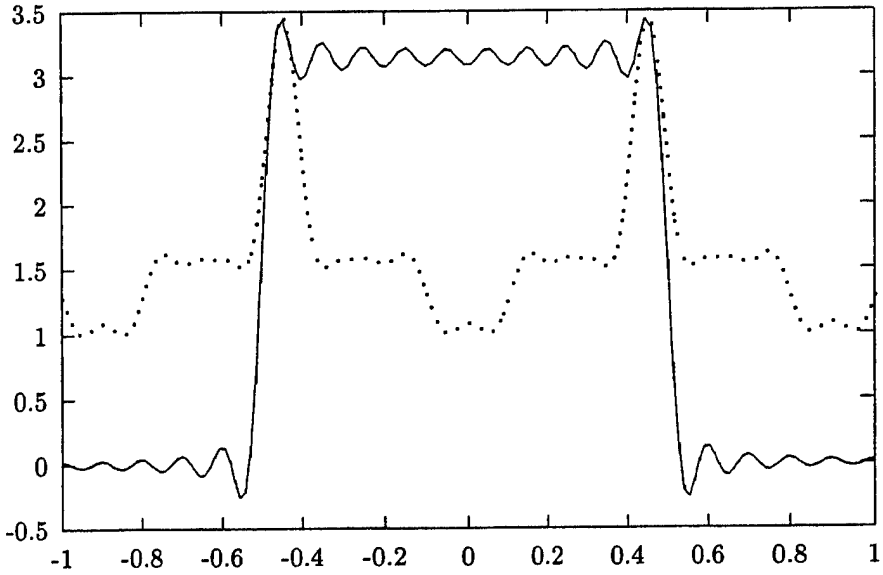


Figure 5b: The optimal approximation to the sector pattern generated by 19 equispaced nodes, as described in Example 5.1

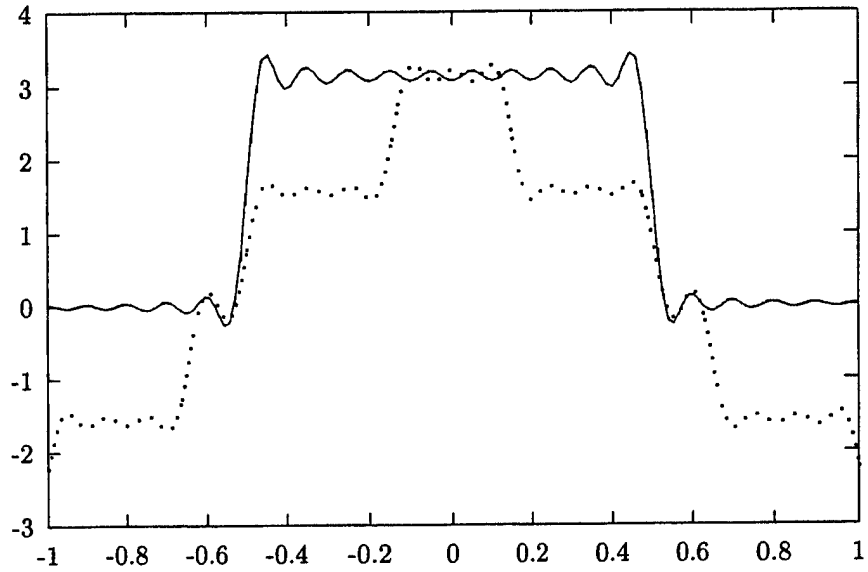


Figure 5c: The optimal approximation to the sector pattern generated by 24 equispaced nodes, as described in Example 5.1

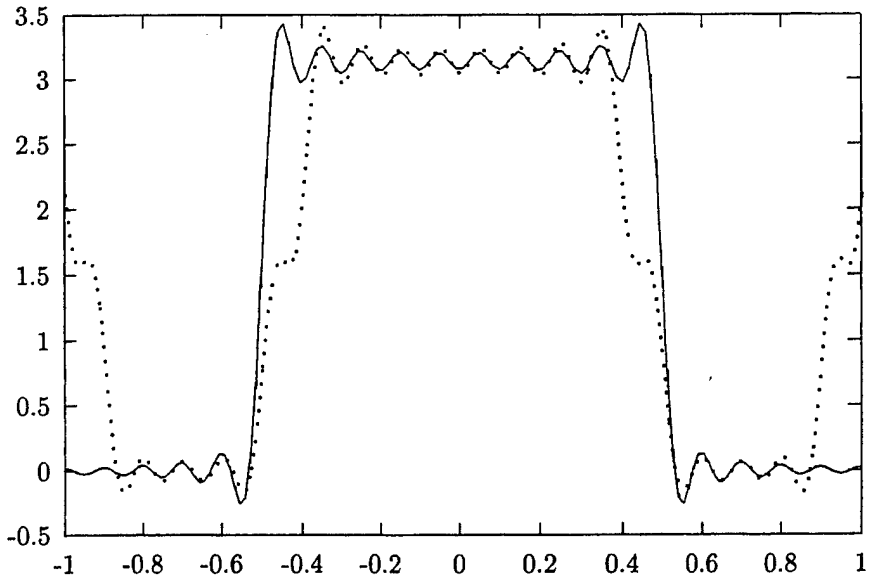


Figure 5d: The optimal approximation to the sector pattern generated by 29 equispaced nodes, as described in Example 5.1

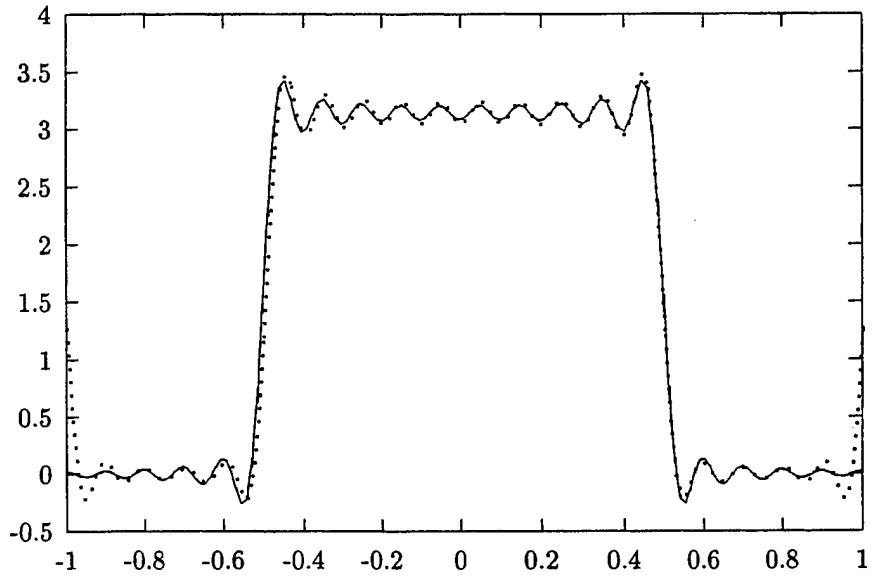


Figure 5e: The optimal approximation to the sector pattern generated by 31 equispaced nodes, as described in Example 5.1

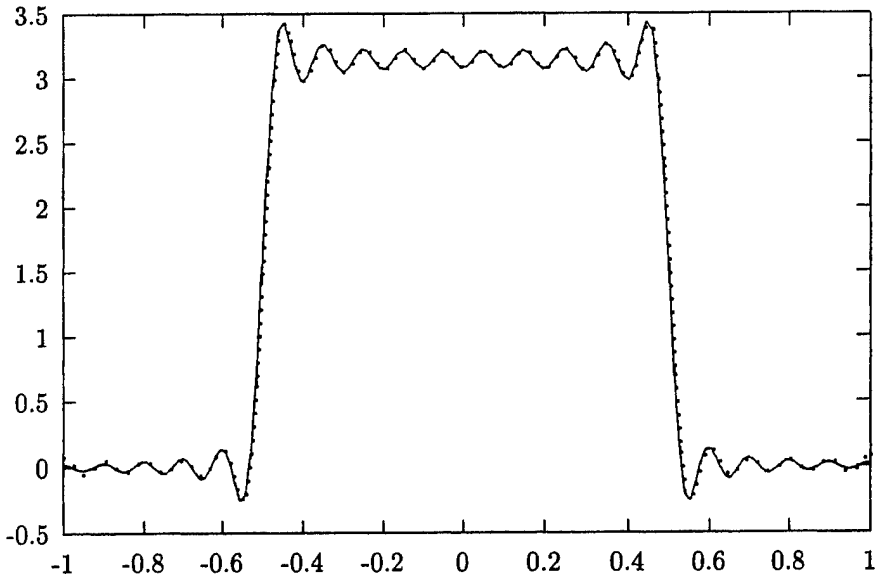


Figure 5f: The optimal approximation to the sector pattern generated by 34 equispaced nodes, as described in Example 5.1

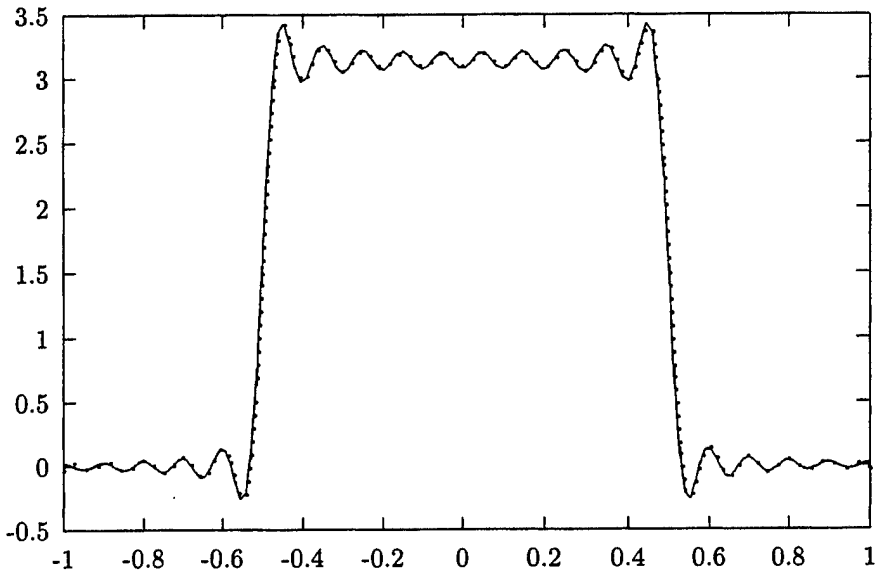


Figure 5g: The optimal approximation to the sector pattern generated by 21 optimal nodes, as described in Example 5.1

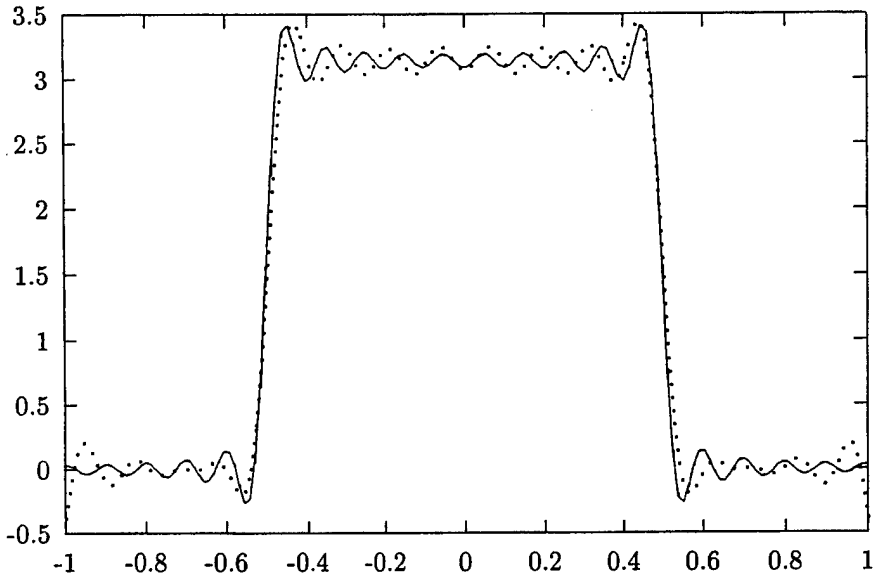


Figure 5h: The optimal approximation to the sector pattern generated by 17 optimal nodes, as described in Example 5.1

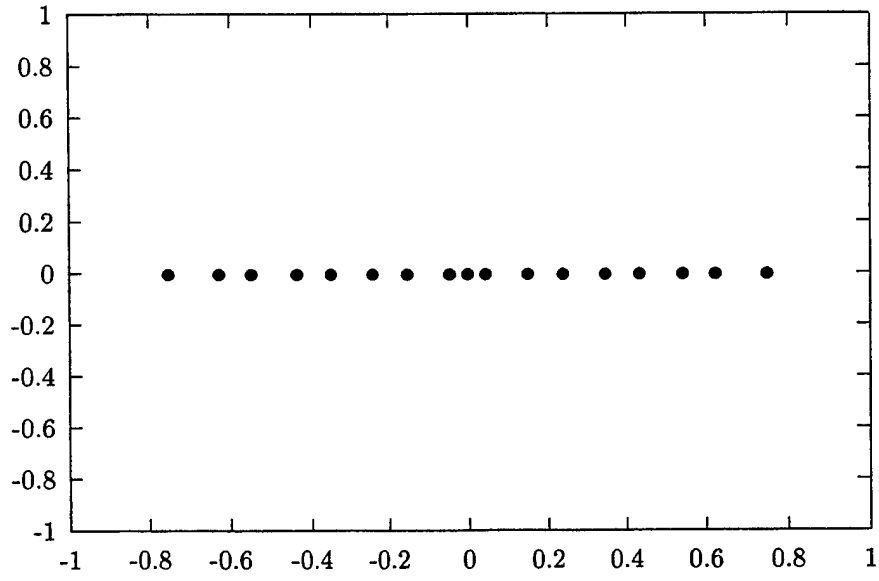


Figure 5i: The distribution of 17 elements creating the pattern depicted in Figure 5h, as described in Example 5.1

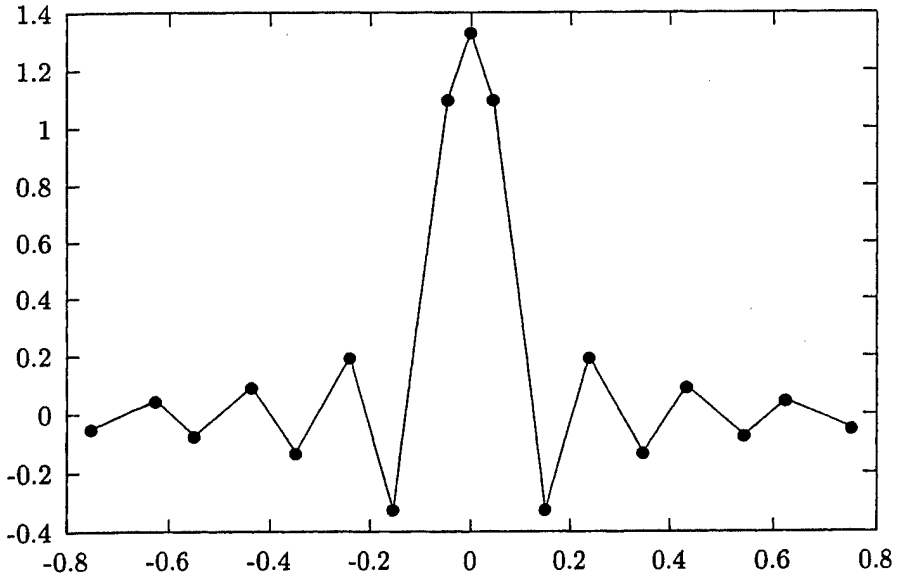


Figure 5j: The values of the sources located at the nodes depicted in Figure 5i and generating the pattern depicted in Figure 5h, as described in Example 5.1

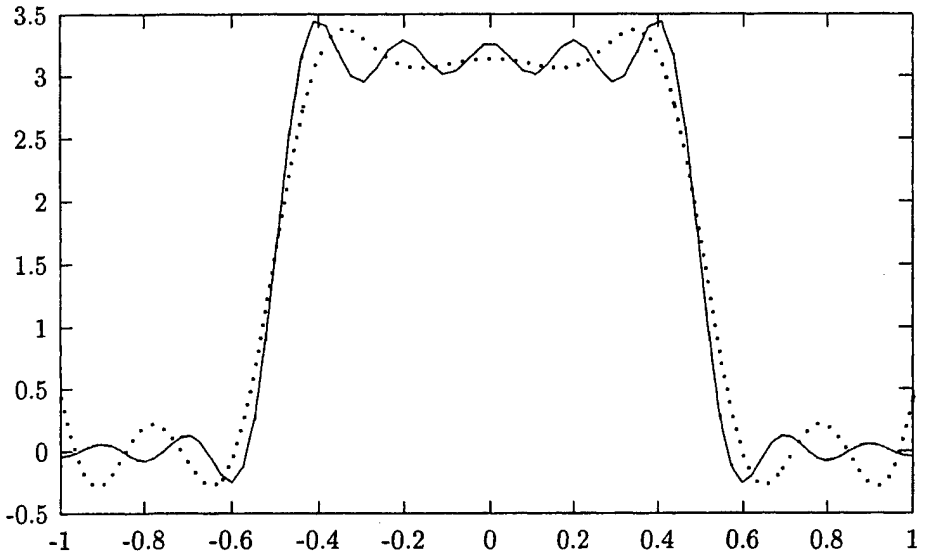


Figure 6: The pattern created by the 9 optimal elements, depicted in Figure 6a as described in Example 5.2

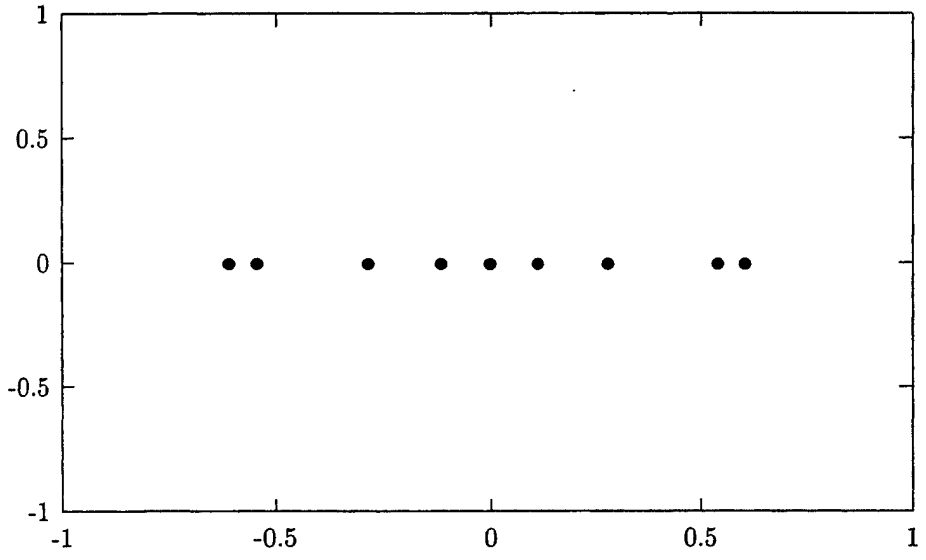


Figure 6a: The distribution of elements creating the pattern depicted in Figure 6, as described in Example 5.2

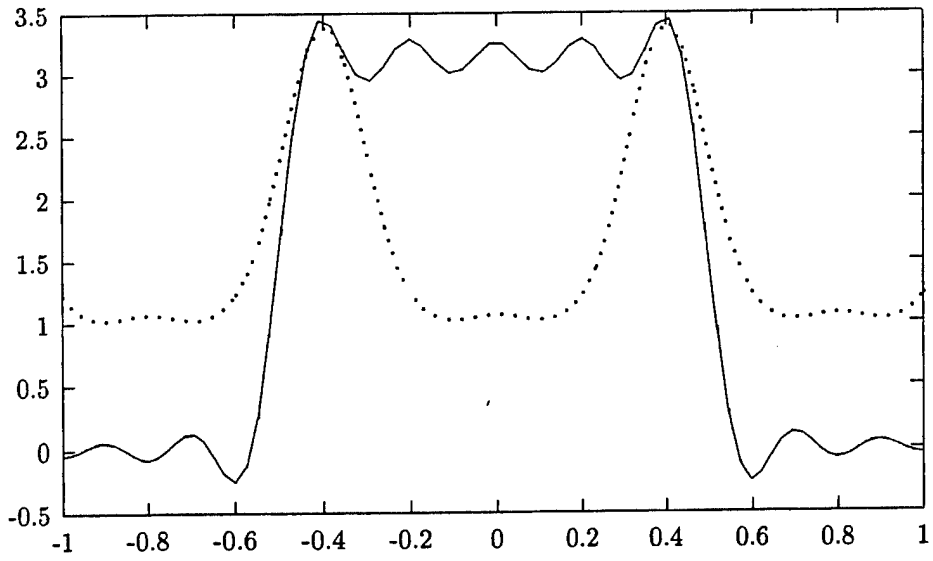


Figure 6b: The optimal approximation to the sector pattern generated by 9 equispaced nodes, as described in Example 5.2

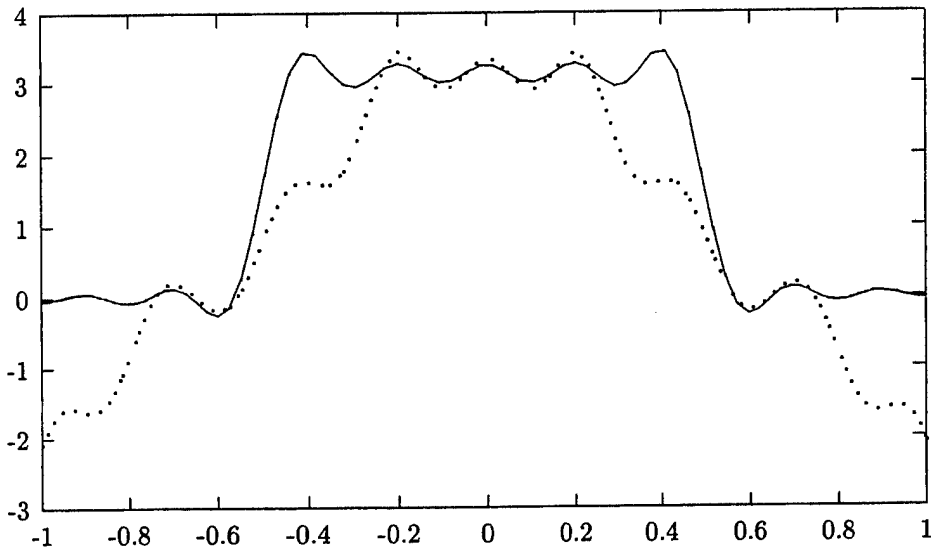


Figure 6c: The optimal approximation to the sector pattern generated by 14 equispaced nodes, as described in Example 5.2

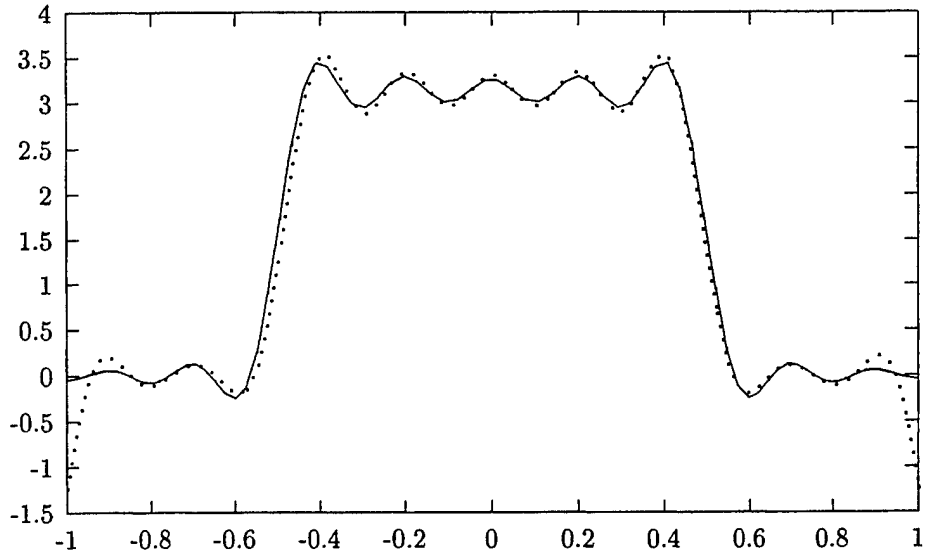


Figure 6d: The optimal approximation to the sector pattern generated by 16 equispaced nodes, as described in Example 5.2

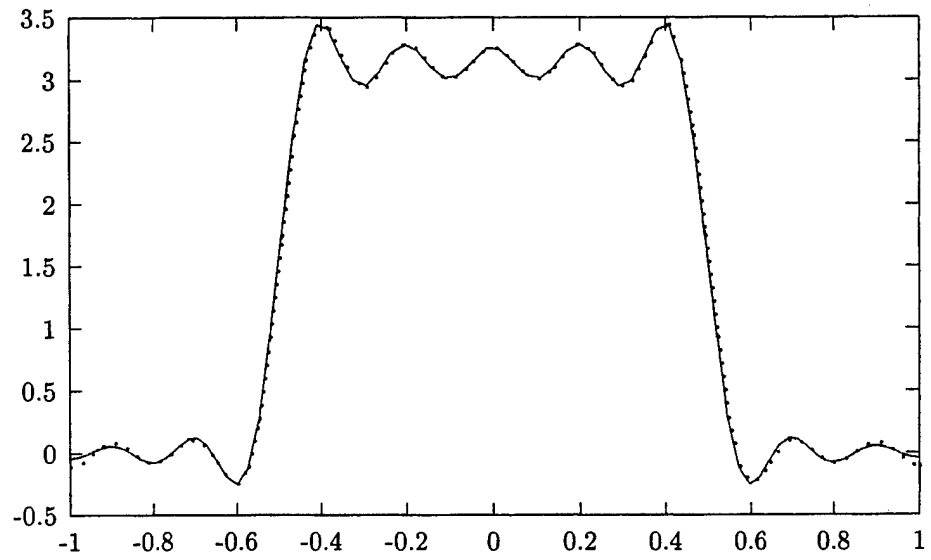


Figure 6e: The optimal approximation to the sector pattern generated by 18 equispaced nodes, as described in Example 5.2

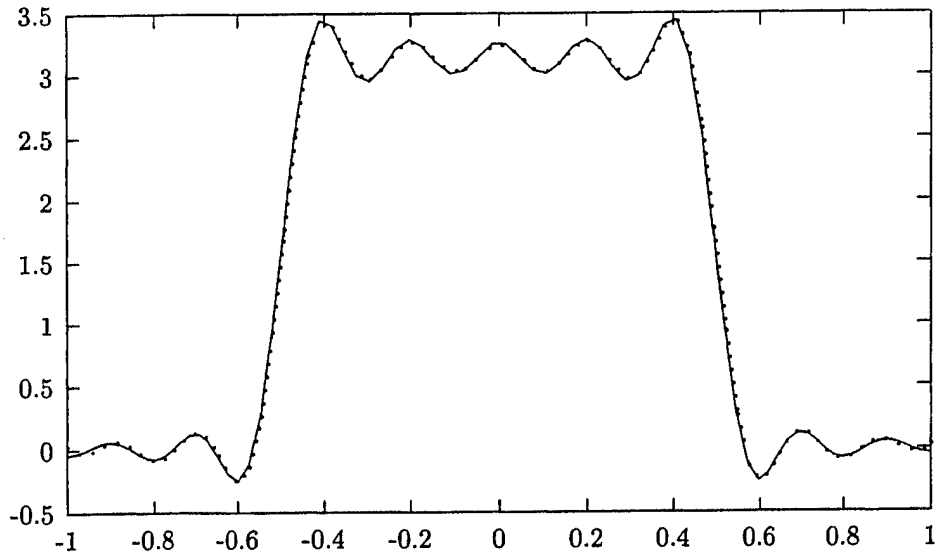


Figure 6f: The pattern created by the 11 optimal elements, in Example 5.2

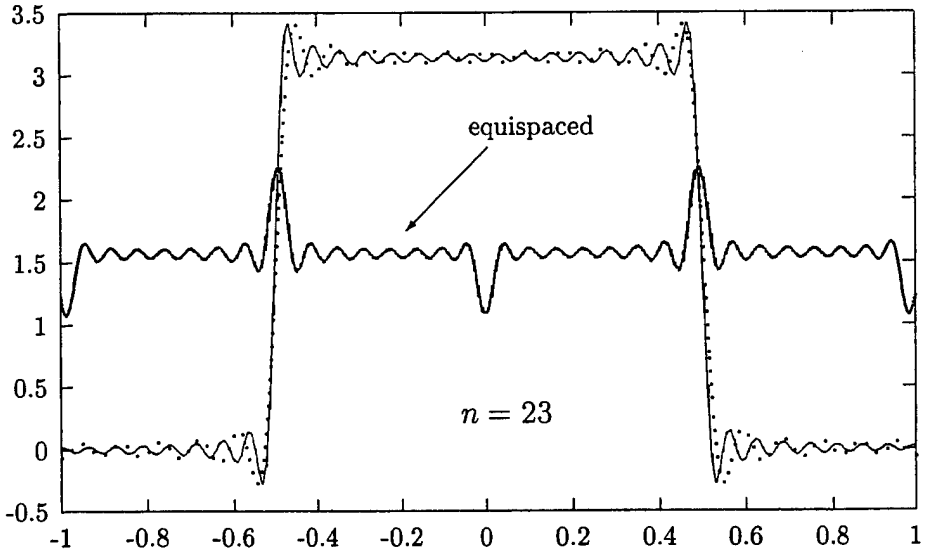


Figure 7a: The approximation to the sector pattern generated by 23 optimal elements, vs. optimal approximation by 23 equispaced nodes, as described in Example 5.3

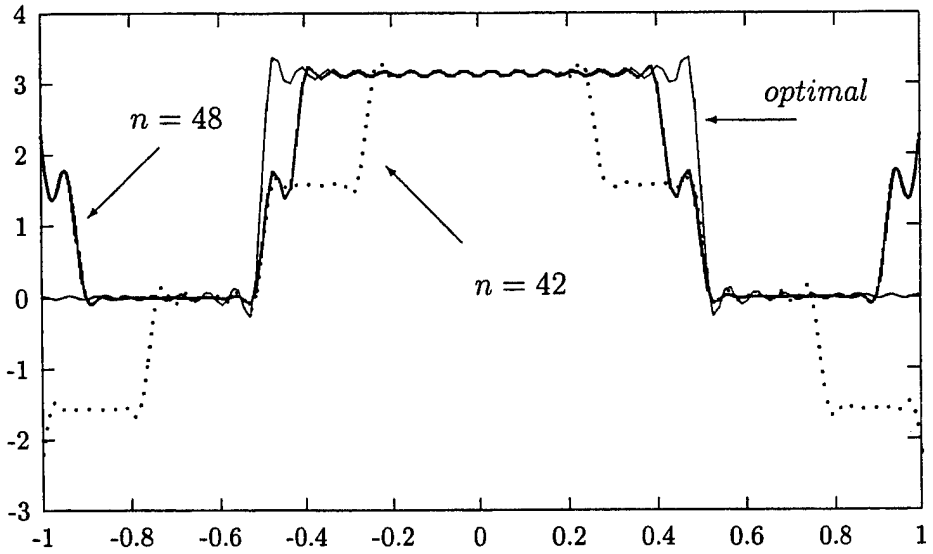


Figure 7b: The optimal approximations to the sector pattern generated by 42 and 48 equispaced nodes, as described in Example 5.3

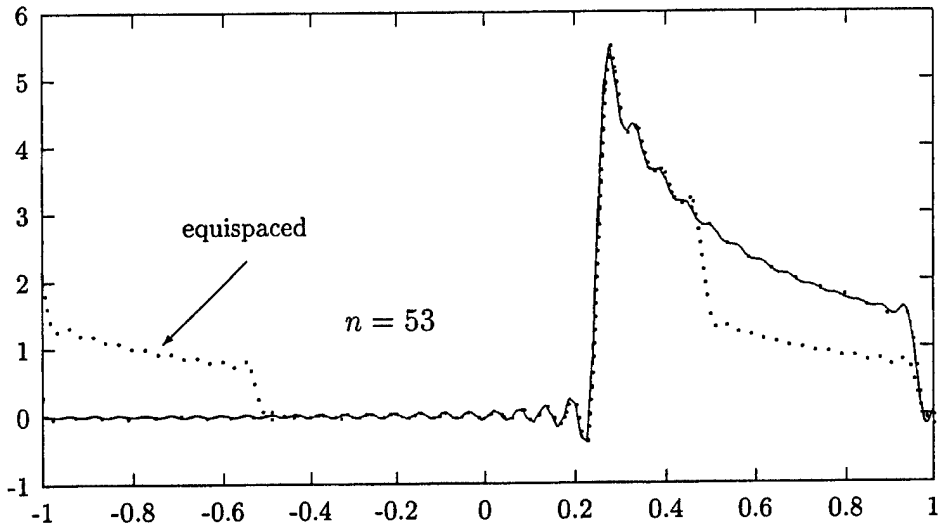


Figure 8a: The approximation to the cosecant pattern generated by 53 optimal elements, vs. optimal approximation by 53 equispaced nodes, as described in Example 5.4

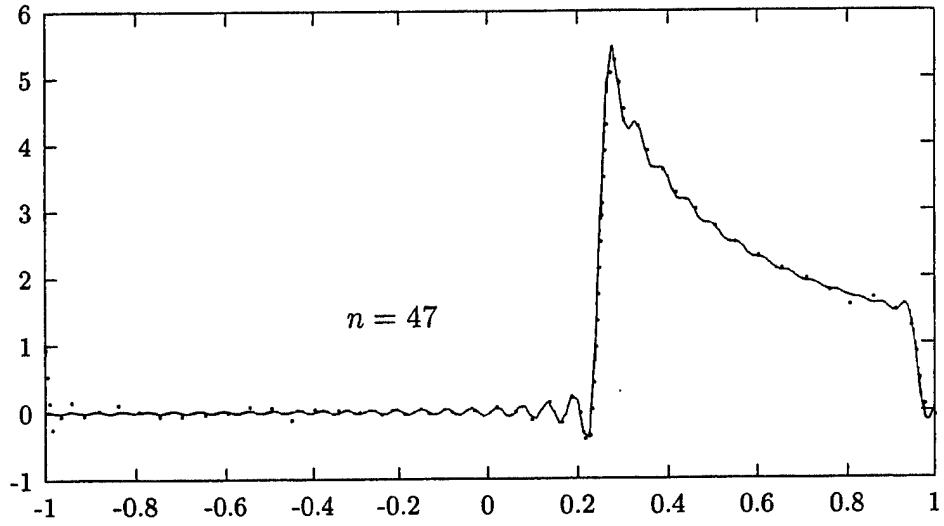


Figure 8a: The approximation to the cosecant pattern generated by 47 optimal elements, as described in Example 5.4

References

[1] M. Abramovitz, I. Stegun, *Handbook of Mathematical Functions*, Applied Math. Series (National Bureau of Standards), Washington, DC, 1964.

[2] H. Cheng, N. Yarvin, V. Rokhlin, *Non-Linear Optimization, Quadrature, and Interpolation*, Yale University Technical Report, YALEU/DCS/RR-1169, 1998, to appear in the SIAM Journal of Non-linear Optimization.

[3] F. GANTMACHER AND M. KREIN, *Oscillation matrices and kernels and small oscillations of mechanical systems*, 2nd ed., Gosudarstv. Izdat. Tehn-Teor. Lit., Moscow, 1950 (Russian).

[4] F. A. Grünbaum, *Toeplitz Matrices Commuting With Tridiagonal Matrices*, J. Linear Alg. and Appl., 40, (1981).

[5] F. A. Grünbaum, *Eigenvectors of a Toeplitz Matrix: Discrete Version of the Prolate Spheroidal Wave Functions*, SIAM J. Alg. Disc. Math., 2(1981).

[6] F. A. Grünbaum, L. Longhi, M. Perlstadt, *Differential Operators Commuting with Finite Convolution Integral Operators: Some Non-Abelian Examples*, SIAM J. Appl. Math. 42(1982).

[7] S. KARLIN, *The Existence of Eigenvalues for Integral Operators*, Trans. Am. Math. Soc. v. 113, pp. 1-17 (1964).

[8] S. KARLIN, AND W. J. STUDDEN, *Tchebycheff Systems with Applications In Analysis And Statistics*, John Wiley (Interscience), New York, 1966.

[9] John D. Kraus, *Antennas*, McGraw-Hill, 1988.

[10] M. G. KREIN, *The Ideas of P. L. Chebyshev and A. A. Markov in the Theory Of Limiting Values Of Integrals*, American Mathematical Society Translations, Ser. 2, Vol. 12, 1959, pp. 1-122.

[11] H.J. Landau, H. Widom, *Eigenvalue Distribution of Time and Frequency Limiting*, Journal of Mathematical Analysis and Applications, 77, 469-481 (1980).

[12] Y.T. Lo, S.W. Lee, editors, *Antenna Handbook, Theory, Applications, and Design*, Van Nostrand Reinhold Company, 1988.

[13] J. MA, V. ROKHLIN, AND S. WANDZURA, *Generalized Gaussian Quadratures For Systems of Arbitrary Functions*, SIAM Journal of Numerical Analysis, v. 33, No. 3, pp. 971-996, 1996.

[14] R.J. Mailloux, *Phased Array Antenna Handbook*, Artech House, 1994.

[15] A. A. MARKOV, *On the limiting values of integrals in connection with interpolation*, Zap. Imp. Akad. Nauk. Fiz.-Mat. Otd. (8) 6 (1898), no.5 (Russian), pp. 146-230 of [16].

[16] A. A. MARKOV, *Selected papers on continued fractions and the theory of functions deviating least from zero*, OGIZ, Moscow-Leningrad, 1948 (Russian).

[17] P.M. Morse, H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, 1953.

[18] D. Rhodes, *The optimum line source for the best mean-square approximation to a given radiation pattern*, IEEE Trans. AP, July 1963.

[19] D. Rhodes, *Synthesis of planar antenna sources*, Clarendon Press, Oxford, 1974.

[20] D. Slepian, H.O. Pollak, *Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty - I*, The Bell System Technical Journal, January 1961.

[21] H.J. Landau, H.O. Pollak, *Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty - II*, The Bell System Technical Journal, January 1961.

[22] H.J. Landau, H.O. Pollak, *Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty - III: The Dimension of Space of Essentially Time- and Band-Limited Signals*, The Bell System Technical Journal, July 1962.

[23] D. Slepian, *Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty - IV: Extensions to Many Dimensions, Generalized Prolate Spheroidal Wave Functions*, The Bell System Technical Journal, November 1964.

[24] D. Slepian, *Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty - V: The Discrete Case*, The Bell System Technical Journal, May-June 1978.

[25] D. Slepian, *Some Comments on Fourier Analysis, Uncertainty, and Modeling* SIAM Review, V. 25, No. 3, July 1983.

[26] W.L. Stutzman, G.A. Thiele, *Antenna Theory and Design*, Wiley, 1998.

[27] T.T. Taylor, *Design of Line-Source Antennas for Narrow Beamwidth and Low Side Lobes*, IEEE Trans. on Antennas and Propagation, AP-3, pp. 16-28, 1955.

[28] N. Yarvin and V. Rokhlin, *Generalized Gaussian Quadratures and Singular Value Decompositions of Integral Operators*, SIAM Journal of Scientific Computing, Vol. 20, No. 2, pp. 699-718 (1998).

A procedure is reported for the compression of rank-deficient matrices. A matrix A of rank k is represented in the form $A = U \circ B \circ V$, where B is a $k \times k$ submatrix of A, and U, V are well-conditioned matrices that each contain a $k \times k$ identity submatrix. This property enables such compression schemes to be used in certain situations where the SVD cannot be used efficiently. Numerical examples are presented.

On the compression of low rank matrices

H. Cheng[†], Z. Gimbutas[†], P.G. Martinsson[‡], V. Rokhlin[‡]

Research Report YALEU/DCS/RR-1251
July 11, 2003

This research was supported in part by the Defense Advanced Research Projects Agency under contract #MDA972-00-1-0033, and by the Office of Naval Research under contract #N00014-01-0364.

[†] MadMax Optics Inc., 3035 Whitney Ave., Hamden CT 06518
[‡] Dept. of Mathematics, Yale University, New Haven CT 06511

Keywords: *Matrix compression, skeletons, model order reduction.*

On the compression of low rank matrices

H. Cheng, Z. Gimbutas, P.G. Martinsson, V. Rokhlin

Abstract: A procedure is reported for the compression of rank-deficient matrices. A matrix A of rank k is represented in the form $A = U \circ B \circ V$, where B is a $k \times k$ submatrix of A, and U, V are well-conditioned matrices that each contain a $k \times k$ identity submatrix. This property enables such compression schemes to be used in certain situations where the SVD cannot be used efficiently. Numerical examples are presented.

1. Introduction

In computational physics (and many other areas), one often encounters matrices whose ranks are (to high precision) much lower than their dimensionalities; even more frequently, one is confronted with matrices posessing large submatrices that are of low rank. An obvious source of such matrices is the potential theory, where discretization of integral equations almost always results in matrices of this type. Such matrices are also encountered in fluid dynamics, numerical simulation of electromagnetic phenomena, structural mechanics, multivariate statistics etc. In such cases, one is tempted to "compress" the matrices in question, so that they could be efficiently applied to arbitrary vectors; compression also facilitates the storage and any other manipulation of such matrices that might be desirable.

At this time, several classes of algorithms exist that use this observation. The so-called Fast Multipole Methods (FMMs) are algorithms for the application of certain classes of matrices to arbitrary vectors; FMMs tend to be extremely efficient, but are only applicable to very narrow classes of operators (see [7]). Another approach to the compression of operators is based on the wavelets and related structures (see, for example, [3, 2]); these schemes exploit the smoothness of the elements of the matrix viewed as a function of their indices, and tend to fail for highly oscillatory operators.

Finally, there is a class of compression schemes that are based purely on linear algebra, and are completely insensitive the the analytical origin of the operator. It consists of the Singular Value Decomposition (SVD), the so-called QR and QLP factorizations [8], and several others. Given an $m \times n$-matrix A of rank $k < \min(m, n)$, the SVD represents A in the form

$$A = U \circ D \circ V, \tag{1.1}$$

with U an $m \times k$, matrix whose columns are orthonormal, V a $k \times n$ matrix whose rows are orthonormal, and D a diagonal matrix whose diagonal elements are positive. The compression provided by the SVD is perfect in terms of accuracy (see, for example, [5]), and has a simple geometric interpretation: it expresses each of the columns of A as a linear combination of the k (orthonormal) columns of U; it also represents the rows of A as linear combinations of (orthonormal) rows of V; and the matrices U, V are chosen in such a manner that the rows of U are images (up to a scaling) under A of the columns of V.

In this paper, we propose a different matrix decomposition. Specifically, we represent the matrix A described above in the form

$$A = \mathcal{U} \circ B \circ \mathcal{V}, \tag{1.2}$$

where B is a $k \times k$-submatrix of A, and the norms of the matrices $\mathcal{U}, \mathcal{V}$ (of dimensionalities $n \times k$, $k \times m$ respectively) are reasonably close to 1 (see Theorem 3 in Section 3 below). Furthermore, each of the matrices $\mathcal{U}, \mathcal{V}$ contains a unity $k \times k$ submatrix.

Like (1.1), the representation (1.2) has a simple geometric interpretation: it expresses each of the columns of A as a linear combination of k selected columns of A, and each of the rows of A as a linear combination of k selected rows of A. This selection defines a $k \times k$ submatrix B of A, and in the resulting system of coordinates, the action of A is represented by the action of its submatrix B.

The representation (1.2) has the advantage that the bases used for the representation of the mapping A consists of the columns and rows of A, while each of the elements of the bases in the representation (1.1) is itself a linear combination of *all* rows (or columns) of the matrix A. In Section 5, we illustrate the advantages of the representation (1.2) by constructing an accelerated direct solver for integral equations of potential theory.

Another advantage of the representation (1.2) is that the numerical procedure for constructing it is considerably less expensive than that for the construction of the SVD (see Section 4), and that the cost of applying (1.2) to an arbitrary vector is

$$(n + m - k) \cdot k, \tag{1.3}$$

vs.

$$(n + m) \cdot k \tag{1.4}$$

for the SVD.

The obvious disadvantage of (1.2) vis-a-vis (1.1) is the fact that the norms of the the matrices $\mathcal{U}, \mathcal{V}$ are somewhat greater than 1, leading to some (though minor) loss of accuracy. Another disadvantage of the proposed factorization is its non-uniqueness; in this respect it is similar to the pivoted QR factorization.

Remark 1. In (1.2), the submatrix B of the matrix A is defined as the intersection of k columns with k rows. Denoting the sequence numbers of the rows by $i_1, i_2, \ldots, i_k$ and the sequence numbers of the columns by $j_1, j_2, \ldots, j_k$, we will be referring to the submatrix B of A as the skeleton of A, to the $k \times n$ matrix consisting of the rows of A numbered $i_1, i_2, \ldots, i_k$ as the row skeleton of A, and to the $m \times k$ matrix consisting of the columns of A numbered $j_1, j_2, \ldots, j_k$ as the column skeleton of A.

The structure of this paper is as follows. Section 2 below summarizes several facts from numerical linear algebra to be used in the remainder of the paper. In Section 3, we prove the existence of a stable factorization of the form (1.2). In Section 4, we describe a reasonably efficient numerical algorithm for constructing such a factorization. In Section 5, we illustrate how the geometric properties of the factorization (1.2) can be utilized to construct an accelerated direct solver for integral equations of potential theory. In Section 6, we present the results of numerical experiments with the direct solver. Finally, Section 7 contains a discussion of other possible applications of the techniques of this paper.

2. Preliminaries

In this section we introduce our notation and summarize several facts from numerical linear algebra; these can all be found in [1].

Throughout the paper, we use upper case letters for matrices and lower case letters for vectors and scalars. We reserve Q for matrices that have orthonormal columns and P for permutation matrices. The canonical unit vectors in $\mathbb{C}^n$ are denoted by e_j. Given a matrix X, we let X^* denote its adjoint (the complex conjugate transpose), $\sigma_k(X)$ its k-th singular value, $||X||_2$ its l^2-norm and $||X||_{\mathrm{F}}$ its Frobenius norm. Finally, given matrices A, B, C and D we let

$$[A|\,B]\,, \qquad \left[\frac{A}{C}\right], \qquad \text{and} \qquad \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right], \tag{2.1}$$

denote larger matrices obtained by stringing the blocks A, B, C and D together.

The first result that we present asserts that given any matrix A, it is possible to reorder its columns to form a matrix AP, where P is a permutation matrix, with the following property: When AP is factorized into an orthonormal matrix Q and an upper triangular matrix R, so that $AP = QR$, then the singular values of the leading $k \times k$ submatrix of R are reasonably good approximations of the first k singular values of A. The theorem also says that the first k columns of AP form a well-conditioned basis for the column space of A to within accuracy $\sigma_{k+1}(A)$.

Theorem 1. [Gu & Eisenstat] *Suppose that A is an $m \times n$ matrix, $l = \min(m,n)$, and k is an integer such that $1 \leq k \leq l$. Then there exists a factorization*

$$AP = QR, \tag{2.2}$$

where P is an $n \times n$ permutation matrix, Q is an $m \times l$ matrix with orthonormal columns, and R is an $l \times n$ upper triangular matrix. Furthermore, splitting Q and R,

$$Q = \left[\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right], \qquad \text{and} \qquad R = \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & R_{22} \end{array}\right], \tag{2.3}$$

in such a fashion that Q_{11} and R_{11} are of size $k \times k$, Q_{21} is $(m-k) \times k$, Q_{12} is $k \times (l-k)$, Q_{22} is $(m-k) \times (l-k)$, R_{12} is $k \times (n-k)$ and R_{22} is $(l-k) \times (n-k)$, results in the following inequalities:

$$\sigma_k(R_{11}) \geq \sigma_k(A) \frac{1}{\sqrt{1 + k(n-k)}}, \tag{2.4}$$

$$\sigma_1(R_{22}) \leq \sigma_{k+1}(A) \sqrt{1 + k(n-k)}, \tag{2.5}$$

and

$$||R_{11}^{-1} R_{12}||_{\mathrm{F}} \leq \sqrt{k(n-k)}. \tag{2.6}$$

Remark 2. In this paper we do not use the full power of Theorem 1 since we are only concerned with the case of very small $\varepsilon = \sigma_{k+1}(A)$. In this case, the inequality (2.5) implies that A can be well approximated by a low-rank matrix. In particular, (2.5) implies that

$$||A - \left[\frac{Q_{11}}{Q_{21}}\right] [R_{11}|\,R_{12}]\, P^*||_2 \leq \varepsilon \sqrt{1 + k(n-k)}. \tag{2.7}$$

Furthermore, the inequality (2.6) in this case implies that the first k columns of AP form a well-conditioned basis for the entire column space of A (within accuracy ε).

While Theorem 1 asserts the existence of a factorization (2.2) with the properties (2.4), (2.5), (2.6), it says nothing about the cost of constructing such a factorization numerically. The following theorem asserts that a factorization that satisfies bounds that are weaker than (2.4), (2.5), (2.6) by a factor of $\sqrt{n}$ can be computed in $O(mn^2)$ operations.

Theorem 2. **[Gu & Eisenstat]** *Given an $m \times n$ matrix A, a factorization of the form (2.2) that instead of (2.4), (2.5) and (2.6) satisfies the inequalities*

$$\sigma_k(R_{11}) \geq \frac{1}{\sqrt{1+nk(n-k)}}\sigma_k(A), \tag{2.8}$$

$$\sigma_1(R_{22}) \leq \sqrt{1+nk(n-k)}\sigma_{k+1}(A), \tag{2.9}$$

and

$$||R_{11}^{-1}R_{12}||_{\mathrm{F}} \leq \sqrt{nk(n-k)}, \tag{2.10}$$

can be computed in $O(mn^2)$ operations.

3. Analytical apparatus

In this section we prove that the factorization (1.2) exists by applying Theorem 1 to both the columns and the rows of the matrix A. Theorem 2 then guarantees that the factorization can be computed efficiently.

The following theorem is the principal analytic tool of this paper.

Theorem 3. *Suppose that A is an $m \times n$ matrix and let k be such that $1 \leq k \leq \min(m,n)$. Then there exists a factorization*

$$A = P_{\mathrm{L}} \begin{bmatrix} I \\ S \end{bmatrix} A_{\mathrm{S}} \begin{bmatrix} I \mid T \end{bmatrix} P_{\mathrm{R}}^{*} + X, \tag{3.1}$$

where $I \in \mathbb{C}^{k\times k}$ is the identity matrix, P_{L} and P_{R} are permutation matrices, and A_{S} is the top left $k \times k$ submatrix of $P_{\mathrm{L}}^{*}AP_{\mathrm{R}}$. In (3.1), the matrices $S \in \mathbb{C}^{(m-k)\times k}$ and $T \in \mathbb{C}^{k\times(n-k)}$ satisfy the inequalities

$$||S||_{\mathrm{F}} \leq \sqrt{k(m-k)}, \qquad \text{and} \qquad ||T||_{\mathrm{F}} \leq \sqrt{k(n-k)}, \tag{3.2}$$

and the matrix X is small if the $(k+1)$-th singular value of A is small,

$$||X||_2 \leq \sigma_{k+1}(A)\sqrt{1+k(\min(m,n)-k)}. \tag{3.3}$$

Proof: The proof consists of two steps. First Theorem 1 is invoked to assert the existence of k columns of A that form a well-conditioned basis for the column space within accuracy $\sigma_{k+1}(A)$; these are collected in the $m \times k$ matrix A_{CS}. Then Theorem 1 is invoked again to prove that k of the rows of A_{CS} form a well-conditioned basis for its row space. Without loss of generality, we assume that $m \geq n$ and that $\sigma_k(A) \neq 0$.

For the first step we factor A into matrices Q and R as specified by Theorem 1, letting P_{R} denote the permutation matrix. Splitting Q and R into submatrices Q_{ij} and R_{ij} as in (2.3), we reorganize the factorization (2.2) as follows,

$$AP_{\mathrm{R}} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} \begin{bmatrix} R_{11} \mid R_{12} \end{bmatrix} + \begin{bmatrix} Q_{12} \\ Q_{22} \end{bmatrix} \begin{bmatrix} 0 \mid R_{22} \end{bmatrix} = \begin{bmatrix} Q_{11}R_{11} \\ Q_{21}R_{11} \end{bmatrix} \begin{bmatrix} I \mid R_{11}^{-1}R_{12} \end{bmatrix} + \left[\begin{array}{c|c} 0 & Q_{12}R_{22} \\ 0 & Q_{22}R_{22} \end{array}\right]. \tag{3.4}$$

We now define the matrix $T \in \mathbb{C}^{k\times(n-k)}$ via the formula

$$T = R_{11}^{-1}R_{12}; \tag{3.5}$$

T satisfies the inequality (3.2) by virtue of (2.6). We define the matrix $X \in \mathbb{C}^{m\times n}$ via the formula

$$X = \left[\begin{array}{c|c} 0 & Q_{12}R_{22} \\ \hline 0 & Q_{11}R_{22} \end{array}\right] P_{\mathrm{R}}^{\star}, \tag{3.6}$$

which satisfies the inequality (3.3) by virtue of (2.5). Defining the matrix $A_{\mathrm{CS}} \in \mathbb{C}^{m\times k}$ by

$$A_{\mathrm{CS}} = \left[\begin{array}{c} Q_{11}R_{11} \\ \hline Q_{21}R_{11} \end{array}\right], \tag{3.7}$$

we reduce equation (3.4) to the form

$$AP_{\mathrm{R}} = A_{\mathrm{CS}}\left[I \middle| T\right] + XP_{\mathrm{R}}. \tag{3.8}$$

An obvious interpretation of (3.8) is that A_{CS} consists of the first k columns of the matrix AP_{R} (since the corresponding columns of XP_{R} are identically zero).

The second step of the proof is to find k rows of A_{CS} forming a well-conditioned basis for its row-space. To this end, we factor the transpose of A_{CS} as specified by Theorem 1,

$$A_{\mathrm{CS}}^{\star}P_{\mathrm{L}} = \tilde{Q}\left[\tilde{R}_{11}\middle|\,\tilde{R}_{12}\right]. \tag{3.9}$$

Transposing (3.9) and rearranging the terms we have

$$P_{\mathrm{L}}^{\star}A_{\mathrm{CS}} = \left[\begin{array}{c} \tilde{R}_{11}^{\star} \\ \hline \tilde{R}_{12}^{\star} \end{array}\right]\tilde{Q}^{\star} = \left[\begin{array}{c} I \\ \hline \tilde{R}_{12}^{\star}(\tilde{R}_{11}^{\star})^{-1} \end{array}\right]\tilde{R}_{11}^{\star}\tilde{Q}^{\star}. \tag{3.10}$$

Multiplying (3.8) by $P_{\mathrm{L}}^{\star}$ and using (3.10) to substitute for $P_{\mathrm{L}}^{\star}A_{\mathrm{CS}}$ we obtain

$$P_{\mathrm{L}}^{\star}AP_{\mathrm{R}} = \left[\begin{array}{c} I \\ \hline \tilde{R}_{12}^{\star}(\tilde{R}_{11}^{\star})^{-1} \end{array}\right]\tilde{R}_{11}^{\star}\tilde{Q}^{\star}\left[I\middle|\,T\right] + P_{\mathrm{L}}^{\star}XP_{\mathrm{R}}. \tag{3.11}$$

We now convert (3.11) into (3.1) by defining the matrices $A_{\mathrm{S}} \in \mathbb{C}^{k\times k}$ and $S \in \mathbb{C}^{(n-k)\times k}$ via the formulæ

$$A_{\mathrm{S}} = \tilde{R}_{11}^{\star}\tilde{Q}^{\star}, \qquad \text{and} \qquad S = \tilde{R}_{12}^{\star}(\tilde{R}_{11}^{\star})^{-1}, \tag{3.12}$$

respectively. □

Remark 3. While the definition (3.5) serves its purpose within the proof of Theorem 3, it is somewhat misleading. Indeed, it is more reasonable to define T as a solution of the equation

$$\|R_{11}T - R_{12}\|_2 \leq \sigma_{k+1}(A)\sqrt{1 + k(n-k)}. \tag{3.13}$$

When the solution is non-unique we chose a solution that minimizes $\|T\|_{\mathrm{F}}$. From the numerical point of view, the definition (3.13) is much preferable to (3.5) since it is almost invariably the case that R_{11} is highly ill-conditioned, if not outright singular.

Introducing the notation

$$A_{\mathrm{CS}} = P_{\mathrm{L}}\left[\begin{array}{c} I \\ \hline S \end{array}\right]A_{\mathrm{S}} \in \mathbb{C}^{n\times k}, \qquad \text{and} \qquad A_{\mathrm{RS}} = A_{\mathrm{S}}\left[I\middle|\,T\right]P_{\mathrm{R}} \in \mathbb{C}^{k\times m}, \tag{3.14}$$

we observe that under the conditions of Theorem 3, the factorization (3.1) can be rewritten in the forms

$$A = A_{\mathrm{CS}}\left[I\middle|\,T\right]P_{\mathrm{R}}^{\star} + X, \tag{3.15}$$

and

$$A = P_{\mathrm{L}} \left[\frac{I}{S}\right] A_{\mathrm{RS}} + X. \tag{3.16}$$

The matrix A_{CS} consists of k of the columns of A, while A_{RS} consists of k of the rows. We refer to A_{S} as the skeleton of A, and to A_{CS} and A_{RS} as the column and row skeletons, respectively.

Remark 4. While Theorem 3 guarantees the existence of a well-conditioned factorization of the form (3.1), it says nothing about the cost of obtaining such a factorization. However, it follows immediately from Theorem 2 that a factorization (3.1) with the matrices S, T, and X satisfying the weaker bounds

$$\|S\|_2 \leq \sqrt{nk(m-k)}, \qquad \text{and} \qquad \|T\|_2 \leq \sqrt{nk(n-k)}, \tag{3.17}$$

and, with $l = \min(m, n)$,

$$\|X\|_2 \leq \sqrt{1 + lk(l-k)}\,\sigma_{k+1}(A), \tag{3.18}$$

can be constructed at the cost $O(mnl)$.

Observation 1. The relations (3.1), (3.15), (3.16) have simple geometric interpretations. Specifically, (3.15) asserts that for a matrix A of rank k, it is possible to select k columns that form a well-conditioned basis of the entire column space. Let $j_1, \ldots, j_k \in \{1, \ldots, n\}$ denote the indices of those columns and let $X_k = \mathrm{span}(e_{j_1}, \ldots, e_{j_k}) \subseteq \mathbb{C}^n$ (thus, X_k is the space of vectors whose only non-zero coordinates are $x_{j_1}, \ldots, x_{j_k}$). According to Theorem 3, there exists an operator

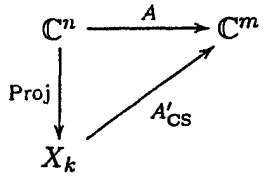
$$\mathrm{Proj} : \mathbb{C}^n \to X_k, \tag{3.19}$$

defined by the formula

$$\mathrm{Proj} = P_{\mathrm{R}} \left[\begin{array}{c|c} I & T \\ \hline 0 & \end{array}\right] P_{\mathrm{R}}^{\star}, \tag{3.20}$$

such that the diagram

(3.21)



is commutative. Here, A'_{CS} is the $m \times n$ matrix formed by setting all columns of A except $j_1, \ldots, j_k$ to zero. Furthermore, $\sigma_1(\mathrm{Proj})/\sigma_k(\mathrm{Proj}) \leq \sqrt{1 + k(n-k)}$. Similarly, equation (3.16) asserts the existence of k rows, say with indices $i_1, \ldots, i_k \in \{1, \ldots, m\}$, that form a well-conditioned basis for the entire row-space. Setting $Y_k = \mathrm{span}(e_{i_1}, \ldots, e_{i_k}) \subseteq \mathbb{C}^m$, there exists an operator

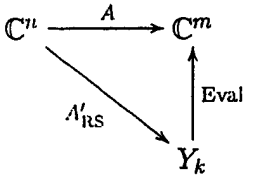
$$\mathrm{Eval} : Y_k \to \mathbb{C}^m, \tag{3.22}$$

defined by

$$\mathrm{Eval} = P_{\mathrm{L}} \left[\begin{array}{c|c} \dfrac{I}{S} & 0 \end{array}\right] P_{\mathrm{L}}^{\star}. \tag{3.23}$$

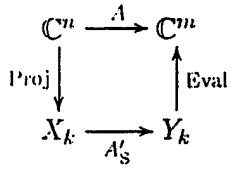
such that the diagram

(3.24)



is commutative. Here, A'_{RS} is the $m \times n$ matrix formed by setting all rows of A except $i_1, \ldots, i_k$ to zero. Furthermore, $\sigma_1(\mathrm{Eval})/\sigma_k(\mathrm{Eval}) \leq \sqrt{1 + k(m-k)}$. Finally, the geometric interpretation of (3.1) is the combination of the diagrams (3.21) and (3.24),

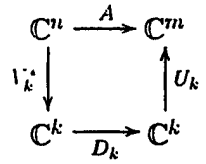
(3.25)



Here, A'_{S} is the $m \times n$ matrix formed by setting all entries of A, except those at the intersection of the rows $i_1, \ldots, i_k$ with the columns $j_1, \ldots, j_k$, to zero.

As a comparison, we consider the diagram

(3.26)



obtained when the SVD is used to compress the matrix $A \in \mathbb{C}^{m \times n}$. Here, D_k is the $k \times k$ diagonal matrix formed by the k largest singular values of A, and V_k and U_k are column matrices containing the corresponding right and left singular vectors, respectively. The factorization (3.25) has the advantage over (3.26) that the mappings Proj and Eval leave k of the coordinates invariant. This is gained at the price of non-orthonormality of these mappings.

4. Numerical apparatus

In this section, we present a simple and reasonably efficient procedure for computing the factorization (3.1). It has been extensively tested and consistently produces factorizations that satisfy the bounds (3.17). While there exist matrices for which this simple approach will not work well, they appear to be exceedingly rare.

Given an $m \times n$ matrix A, the first step (out of four) is to apply the pivoted Gram-Schmidt process to its columns. The process is halted when the column space has been exhausted to a preset accuracy ε, leaving a factorization

$$AP_{\mathrm{R}} = Q \left[R_{11} \middle| R_{12} \right], \tag{4.1}$$

where $P_{\mathrm{R}} \in \mathbb{C}^{n \times n}$ is a permutation matrix, $Q \in \mathbb{C}^{m \times k}$ has orthonormal columns, $R_{11} \in \mathbb{C}^{k \times k}$ is upper triangular, and $R_{12} \in \mathbb{C}^{k \times (n-k)}$.

The second step is to find a matrix $T \in \mathbb{C}^{k \times (n-k)}$ that solves the equation

$$R_{11} T = R_{12} \tag{4.2}$$

to within accuracy ε. When R_{11} is ill-conditioned, there is a large set of solutions; we pick one for which $||T||_{\mathrm{F}}$ is small.

Letting $A_{\mathrm{CS}} \in \mathbb{C}^{m\times k}$ denote the matrix formed by the first k columns of AP_{R}, we now have a factorization

$$A = A_{\mathrm{CS}} \left[I \middle| T\right] P_{\mathrm{R}}^{*}. \tag{4.3}$$

The third and the fourth steps are entirely analogous to the first and the second, but are concerned with finding k rows of A_{CS} that form a basis for its row-space. They result in a factorization

$$A_{\mathrm{CS}} = P_{\mathrm{L}} \begin{bmatrix} I \\ \hline S \end{bmatrix} A_{\mathrm{S}}. \tag{4.4}$$

The desired factorization is now obtained by inserting (4.4) into (4.3):

$$A = P_{\mathrm{L}} \begin{bmatrix} I \\ \hline S \end{bmatrix} A_{\mathrm{S}} \left[I \middle| T\right] P_{\mathrm{R}}^{*}. \tag{4.5}$$

For this technique to be successful, it is crucially important that the Gram-Schmidt factorization be performed accurately. Modified Gram-Schmidt or the method using Householder reflectors are not accurate enough. Instead, we use a technique that is based on modified Gram-Schmidt, but that at each step re-orthogonalizes the vector chosen to add to the basis before adding it. In exact arithmetic, this step would be superfluous, but in the presence of round-off error it greatly increases the quality of the factorization generated, see *e.g.* [6].

5. Application: An accelerated direct solver for integral equations

In this section we use the matrix compression technique presented in Section 3 to construct an accelerated direct solver for boundary integral equations with non-oscillatory kernels. Upon discretization, such equations lead to dense systems of linear equations, and iterative methods combined with fast matrix-vector multiplication techniques are commonly used to obtain the solution. Many such fast multiplication techniques take advantage of the fact that the off-diagonal blocks of the discrete system typically have low rank. Employing the matrix compression techniques presented in Section 3, we use this low-rank property to accelerate direct, rather than iterative, solution techniques. The method uses no machinery beyond what is described in Section 3 and is applicable to most integral equations involving non-oscillatory kernels.

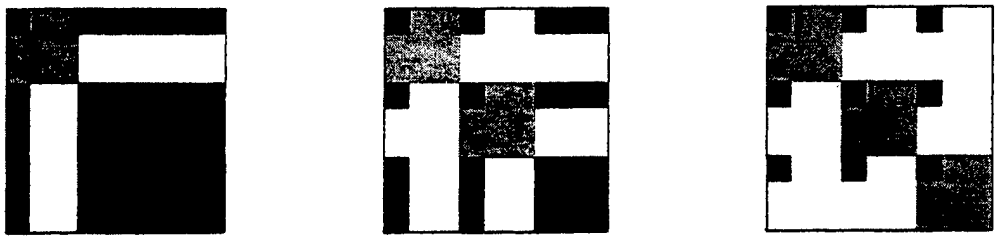
For concreteness, we consider the equation

$$u(x) + \int_{\Gamma} K(x,y)u(y)\,dy = f(x), \qquad \text{for } x \in \Gamma, \tag{5.1}$$

where Γ is some contour and $K(x,y)$ is a non-oscillatory kernel. The function u represents an unknown "charge" distribution on Γ that is to be determined from the given function f. The method that we present works for almost any contour but for simplicity, we will assume that the contour consists of p disjoint pieces, $\Gamma = \Gamma_1 + \cdots + \Gamma_p$, where all pieces have similar size (an example is given in Fig. 3). In fact, to simplify the formulas, we will for the most part set $p = 3$.

Discretizing each contour Γ_i using n points, the equation (5.1) takes the form

$$\begin{bmatrix} M^{(1,1)} & M^{(1,2)} & M^{(1,3)} \\ \hline M^{(2,1)} & M^{(2,2)} & M^{(2,3)} \\ \hline M^{(3,1)} & M^{(3,2)} & M^{(3,3)} \end{bmatrix} \begin{bmatrix} u^{(1)} \\ \hline u^{(2)} \\ \hline u^{(3)} \end{bmatrix} = \begin{bmatrix} f^{(1)} \\ \hline f^{(2)} \\ \hline f^{(3)} \end{bmatrix}, \tag{5.2}$$



(a) (b) (c)

FIGURE 1. Zeros are introduced into the matrix in three steps: (a) interaction between Γ_1 and the other contours is compressed, (b) interaction with Γ_2 is compressed, (c) interaction with Γ_3 is compressed. The small black blocks are of size $k \times k$ and consist of entries that have not been changed beyond permutations, grey blocks refer to updated parts and white blocks are all zero entries.

where $u^{(i)} \in \mathbb{C}^n$ and $f^{(i)} \in \mathbb{C}^n$ are discrete representations of the unknown boundary charge distribution and the right hand side associated with Γ_i, and $M^{(i,j)} \in \mathbb{C}^{n \times n}$ is a dense matrix representing the evaluation of a potential on Γ_i caused by a charge distribution on Γ_j.

The interaction between Γ_1 and the rest of the contour is governed by the matrices

$$H^{(1)} = \left[M^{(1,2)} \middle| M^{(1,3)} \right] \in \mathbb{C}^{n \times 2n}, \qquad \text{and} \qquad V^{(1)} = \left[\frac{M^{(2,1)}}{M^{(3,1)}} \right] \in \mathbb{C}^{2n \times n}. \tag{5.3}$$

For non-oscillatory kernels, these matrices are typically highly rank-deficient. We let k denote an upper bound on their ranks (to within some preset level of accuracy ε). By virtue of (3.16), we know that there exist k rows of $H^{(1)}$ which form a well-conditioned basis for all the n rows. In other words, there exists a well-conditioned $n \times n$ matrix $L^{(1)}$ (see Remark 6) such that

$$L^{(1)} H^{(1)} = \left[\frac{H^{(1)}_{\text{RS}}}{Z} \right] + O(\varepsilon), \tag{5.4}$$

where $H^{(1)}_{\text{RS}}$ is a $k \times 2n$ matrix formed by k of the rows of $H^{(1)}$ and Z is the $(n-k) \times 2n$ zero matrix. There similarly exist an $n \times n$ matrix $R^{(1)}$ such that

$$V^{(1)} R^{(1)} = \left[V^{(1)}_{\text{CS}} \middle| Z^* \right] + O(\varepsilon), \tag{5.5}$$

where $V^{(1)}_{\text{CS}}$ is a $2n \times k$ matrix formed by k of the columns of $V^{(1)}$. For simplicity, we will henceforth assume that the off-diagonal blocks have *exact* rank at most k and ignore the error terms.

The relations (5.4) and (5.5) imply that by restructuring equation (5.2) as follows,

$$\left[\begin{array}{c|c|c} L^{(1)} M^{(1,1)} R^{(1)} & L^{(1)} M^{(1,2)} & L^{(1)} M^{(1,3)} \\ \hline M^{(2,1)} R^{(1)} & M^{(2,2)} & M^{(2,3)} \\ \hline M^{(3,1)} R^{(1)} & M^{(3,2)} & M^{(3,3)} \end{array} \right] \left[\begin{array}{c} (R^{(1)})^{-1} u^{(1)} \\ \hline u^{(2)} \\ \hline u^{(3)} \end{array} \right] = \left[\begin{array}{c} L^{(1)} f^{(1)} \\ \hline f^{(2)} \\ \hline f^{(3)} \end{array} \right], \tag{5.6}$$

we introduce large blocks of zeros in the matrix, as shown in Figure 1(a).

Next, we compress the interaction between Γ_2 and the rest of the contour to obtain the matrix structure shown in Fig. 1(b). Repeating the process with Γ_3, we obtain the final structure shown

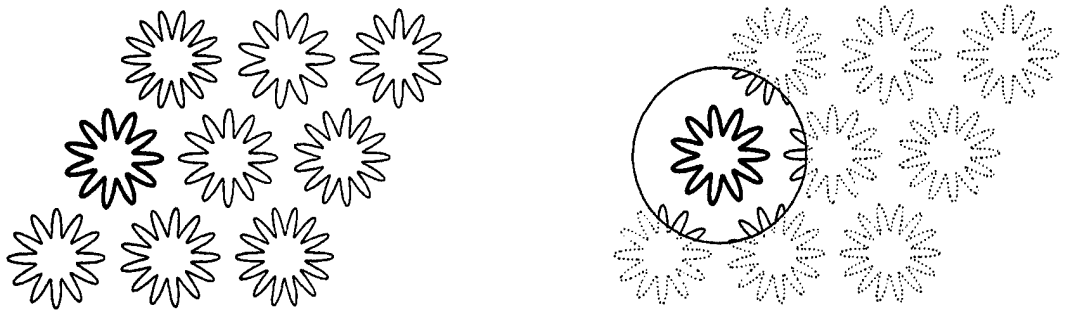


FIGURE 2. In order to determine the $R^{(i)}$ and $L^{(i)}$ that compress the interaction between Γ_i (shown in bold) and the remaining contours, it is sufficient to consider only the interactions between the contours drawn with a solid line in (b).

in Fig. 1(c). At this point, we have constructed matrices $R^{(i)}$ and $L^{(i)}$ and formed the new system

$$\left[\begin{array}{c|c|c} L^{(1)}M^{(1,1)}R^{(1)} & L^{(1)}M^{(1,2)}R^{(2)} & L^{(1)}M^{(1,3)}R^{(3)} \\ \hline L^{(2)}M^{(2,1)}R^{(1)} & L^{(2)}M^{(2,2)}R^{(2)} & L^{(2)}M^{(2,3)}R^{(3)} \\ \hline L^{(3)}M^{(3,1)}R^{(1)} & L^{(3)}M^{(3,2)}R^{(2)} & L^{(3)}M^{(3,3)}R^{(3)} \end{array}\right] \left[\begin{array}{c} (R^{(1)})^{-1}u^{(1)} \\ \hline (R^{(2)})^{-1}u^{(2)} \\ \hline (R^{(3)})^{-1}u^{(3)} \end{array}\right] = \left[\begin{array}{c} L^{(1)}f^{(1)} \\ \hline L^{(2)}f^{(2)} \\ \hline L^{(3)}f^{(3)} \end{array}\right], \tag{5.7}$$

whose matrix is shown in Figure 1(c). We emphasize that the $k \times k$ non-zero parts of the off-diagonal blocks are *submatrices* of the original $n \times n$ off-diagonal blocks. The parts of the matrix that are shown as grey in the figure represent interactions that are internal to each contour. These $n-k$ degrees of freedom per contour can be eliminated by performing a local, $O(n^3)$, operation for each contour. This leaves a dense system of 3×3 blocks, each of size $k \times k$. Thus, we have reduced the problem size by a factor of n/k.

Remark 5. For the algorithm presented above, the compression of the interaction between a fixed contour and its $p-1$ fellows is quite costly since it requires the construction and compression of the large matrices $H^{(i)} \in \mathbb{C}^{n\times(p-1)n}$ and $V^{(i)} \in \mathbb{C}^{(p-1)n\times n}$. In the numerical examples presented below, this step is avoided by constructing matrices $L^{(i)}$ and $R^{(i)}$ that satisfy (5.4) and (5.5) through an entirely local procedure. We illustrate how this is done by considering the contours in Fig. 2(a) and supposing that we want to find the transforms that compress the interaction of the contour Γ_i (drawn with a bold line) with the remaining ones. This can be done by compressing the interaction between Γ_i and an artificial contour Γ_{artif} that surrounds Γ_i (as shown in Fig. 2(b)) combined with the parts of the other contours that penetrate it. This procedure works for any potential problem for which the Green's identities hold. The computational cost for one compression is $O(kn^2)$ rather than the $O(pkn^2)$ cost for constructing and compressing the entire $H^{(i)}$ and $V^{(i)}$.

To sum up: The accelerated solver consists of four steps. For a problem involving p contours, each of which is discretized using n nodes and having off-diagonal blocks of rank at most k, they are:

(1) The off-diagonal blocks are skeletonized and the diagonal $n \times n$ blocks are updated at a cost of $O(pkn^2)$ using the technique described in Remark 5.

(2) The $n - k$ degrees of freedom that represent internal interactions for each contour are eliminated at a cost of $O(pn^3)$.
(3) The reduced $kp \times kp$ system is solved at a cost of $O(k^3p^3)$.
(4) The solution of the original system is reconstructed from the solution of the reduced problem through p local operations at a cost of $O(pn^2)$.

The third step is typically the most expensive one with an asymptotic cost of $t^{(\text{comp})} \sim ck^3p^3$. The cost of a solution of the uncompressed equations is $t^{(\text{uncomp})} \sim cn^3p^3$. Consequently;

$$\text{Speed-up} = \frac{t^{(\text{uncomp})}}{t^{(\text{comp})}} \sim \left(\frac{n}{k}\right)^3.$$

Remark 6. The existence of the matrices $L^{(1)}$ and $R^{(1)}$ are direct consequences of (3.16) and (3.15), respectively. Specifically, substituting $H^{(1)}$ for A in (3.16), we obtain

$$P_{\text{L}}^{\star} H^{(1)} = \begin{bmatrix} I \\ \hline S \end{bmatrix} H_{\text{RS}}^{(1)}, \tag{5.8}$$

where $H_{\text{RS}}^{(1)}$ is the $k \times 2n$ matrix consisting of the top k rows of $P_{\text{L}}^{\star} H^{(1)}$. The relation (5.4) now follows from (5.8) by defining

$$L^{(1)} = \left[\begin{array}{c|c} I & 0 \\ \hline -S & I \end{array}\right] P_{\text{L}}^{\star}. \tag{5.9}$$

We note that the largest and smallest singular values of $L^{(1)}$ satisfy

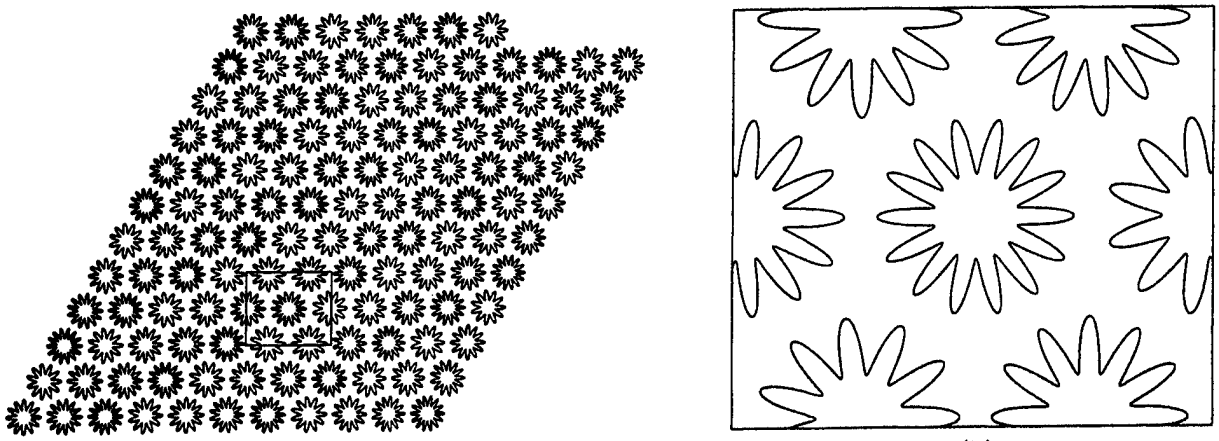
$$\begin{aligned} \sigma_1(L^{(1)}) &\leq \left(1 + \|S\|_{l^2}^2\right)^{1/2}, \\ \sigma_n(L^{(1)}) &\geq \left(1 + \|S\|_{l^2}^2\right)^{-1/2}. \end{aligned} \tag{5.10}$$

Thus $\text{cond}(L^{(1)}) \leq 1 + \|S\|_{l^2}^2$, which is of moderate size according to Theorem 3. The matrix $R^{(1)}$ is similarly constructed by forming the column skeleton of $V^{(1)}$.

Remark 7. Equations (5.4) and (5.5) have simple heuristic interpretations: Equation (5.4) says that it is possible to choose k points on the contour Γ_1 in such a way that when a field generated by charge distributions on the rest of the contour is known at those points, it is possible to extrapolate the field at the remaining points on Γ_1 from those values. Equation (5.5) says that it is possible to choose k points on Γ_1 in such a way that any field on the rest of the contour generated by charges on Γ_1, can be replicated by placing charges only on those k points.

Remark 8. It is sometimes advantageous to choose the same k points when constructing the skeletons of $H^{(i)}$ and $V^{(i)}$. This can be achieved by compressing the two matrices jointly, for instance by forming the row skeleton of $\left[H^{(i)} \middle| (V^{(i)})^{\star}\right]$. In this case $L^{(i)} = (R^{(i)})^{\star}$. When this is done, the compression ratio deteriorates since the singular values of $\left[H^{(i)} \middle| (V^{(i)})^{\star}\right]$ decay slower than those of either $H^{(i)}$ or $V^{(i)}$, as is seen by comparing Figures 4 and 5.

Remark 9. When the solution of equation (5.2) is sought for multiple right-hand sides, the cost of the first solve is $O(mnk)$. Subsequent solves can be preformed using $O(p^2k^2 + pn^2)$ operations rather than $O(p^2n^2)$ for an uncompressed solver.



(a) (b)

FIGURE 3. The contours used for the numerical calculations with $p = 128$. Picture (a) shows the full contour and a box (which is not part of the contour) that indicates the location of the close-up shown in (b).

Remark 10. The direct solver that we have presented has a computational complexity that scales cubically with the problem size N and is thus not a "fast" algorithm. However, by applying the techniques presented recursively, it is possible to reduce the asymptotic complexity to $O(N^{3/2})$, and possibly even $O(N \log N)$. This is a topic of current research.

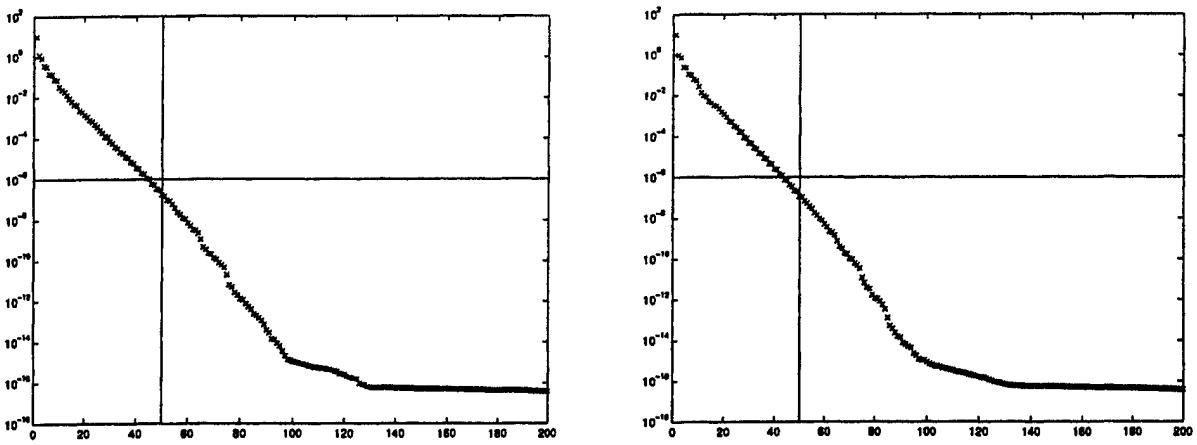
6. NUMERICAL RESULTS

The algorithm described in Section 5 has been computationally tested on the second kind integral equation obtained by discretizing an exterior Dirichlet boundary value problem using the double layer kernel. The contours used consisted of a number of jagged circles arranged in a skewed square as shown in Fig. 3. The number of contours p ranged from 8 to 128. For this problem, $n = 200$ points per contour were required to obtain a relative accuracy of $\varepsilon = 10^{-6}$. We found that to this level of accuracy, no $H^{(i)}$ or $V^{(i)}$ had rank exceeding $k = 50$. As an example, we show in Fig. 4 the singular values of the matrices $H^{(i)}$ and $V^{(i)}$ representing interactions between the highlighted contour in Fig. 2(a) and the remaining ones.

The algorithm described in Section 5 was implemented in FORTRAN and run on a 2.8GHz Pentium IV desktop PC with 512Mb RAM. The CPU times for a range of different problem sizes are presented in Table 1. The data presented supports the following claims for the compressed solver:

- For large problems, the CPU time speed-up approaches the estimated factor of $(n/k)^3 = 64$.
- The reduced memory requirement make large problems amenable to direct solution.

Remark 11. In the interest of simplicity, we forced the program to use the same compression ratio k/n for each contour. In general, it detects the required interaction rank of each contour as its interaction matrices are being compressed and uses different ranks for each contour.



(a) (b)

FIGURE 4. Plots of the singular values of (a) $V^{(i)}$ and (b) $H^{(i)}$ for a discretization of the double layer kernel associated with the Laplace operator on the nine contours depicted in Fig. 2(a). In the example shown, the contours were discretized using $n = 200$ points, giving a relative discretization error of about 10^{-6}. The plots show that to that level of accuracy, the matrices $V^{(i)} \in \mathbb{C}^{1600\times 200}$ and $H^{(i)} \in \mathbb{C}^{200\times 1600}$ have numerical rank less than $k = 50$ (to accuracy 10^{-6}).

p	$t^{(\text{uncomp})}$	$t^{(\text{comp})}$	$t^{(\text{comp})}_{\text{init}}$	$t^{(\text{comp})}_{\text{solve}}$	Error
8	5.6	2.0 (4.6)	1.6 (4.1)	0.05	$8.1 \cdot 10^{-7} (1.4 \cdot 10^{-7})$
16	50	4.1 (16.4)	3.1 (15.5)	0.4	$2.9 \cdot 10^{-6} (2.8 \cdot 10^{-7})$
32	451	13.0 (72.1)	6.4 (65.3)	5.5	$4.4 \cdot 10^{-6} (4.4 \cdot 10^{-7})$
64	*3700*	65 (270)	14 (220)	48	—
128	*30000*	480 (1400)	31 (960)	440	—

TABLE 1. CPU times in seconds for solving (5.2). p is the number of contours. $t^{(\text{uncomp})}$ is the CPU time required to solve the uncompressed equations; the numbers in italics are estimated since these problems did not fit in RAM. $t^{(\text{comp})}$ is the CPU time to solve the equations using the compression method; this time is split between $t^{(\text{comp})}_{\text{init}}$, the time to compress the equations, and $t^{(\text{comp})}_{\text{solve}}$, the time to solve the reduced system of equations. The error is the relative error incurred by the compression measured in the maximum norm when the right hand side is a vector of ones. Throughout the table, the numbers in parenthesis refer to numbers obtained when the technique of Remark 5 is not used.

7. CONCLUSIONS

We have described a "compression" scheme for low-rank matrices. For a matrix A of dimensionality $m \times n$ and rank k, the factorization can be applied to an arbitrary vector for the cost of $(n+m-k)\cdot k$ operations, after a significant initial factorization cost; this is marginally faster than

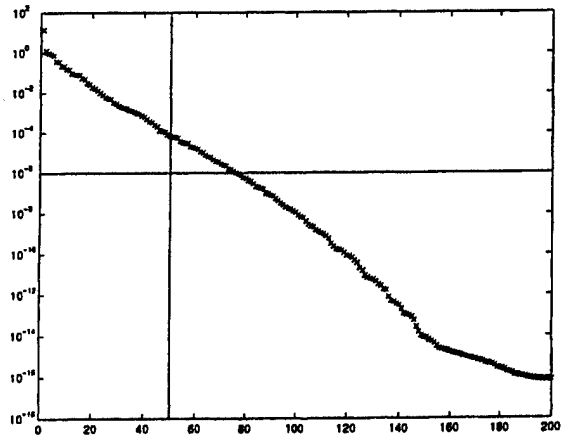


FIGURE 5. Plot of the singular values of $X^{(i)} = \left[H^{(i)} \middle| (V^{(i)})^*\right]$ where $H^{(i)}$ and $V^{(i)}$ are as in Figure 4. The numerical rank of $X^{(i)}$ is approximately 80, which is larger than the individual ranks of $H^{(i)}$ and $V^{(i)}$.

the cost $(n+m) \cdot k$ produced by the SVD. The factorization cost is roughly the same as that for the rank-revealing QR decomposition of A.

A more important advantage of the proposed decomposition is the fact that it expresses all of the columns of A as linear combinations of k appropriately selected columns of A, and all of the rows of A as linear combinations of k appropriately selected rows of A. Since each of the basis vectors (both row and column) produced by the SVD (or any other classical factorizations) is a linear combination of *all* rows (columns) of A, the decomposition we propose is considerably easier to manipulate; we illustrate this point by constructing an accelerated scheme for the direct solution of integral equations of potential theory in the plane.

A related advantage of the proposed decomposition is the fact that one frequently encounters collections of matrices such that the same selection of rows and columns can be used for each matrix to span its row and column space (in other words, there exist fixed P_{L} and P_{R} such that each matrix in the collection has a decomposition (3.1) with small matrices S and T). Once one matrix in such a collection has been factorized, the decomposition of the remaining ones is considerably simplified since the skeleton of the first can be reused. If it should happen that the skeleton of the first matrix that was decomposed is not a good choice for some other matrix, this is easily detected (since then no small matrices S and T can be computed) and the global skeleton can be extended as necessary.

We have constructed several other numerical procedures using the approach described in this paper. In particular, a code has been designed for the (reasonably) rapid solution of scattering problems in the plane based on the direct (as opposed to iterative) solution of the Lippman-Schwinger equation; the scheme utilizes the same idea as that used in [4], and has the same asymptotic CPU time estimate $O(N^{3/2})$ for a square region discretized into N nodes. However, the CPU times obtained by us are a significant improvement on these reported in [4]; the paper reporting this work is in preparation.

It also appears to be possible to utilize the techniques of this paper to construct an order $O(N \log N)$ (or possibly even order order $O(N)$ (!)) scheme for the solution of elliptic PDEs in

both two and three dimensions, provided that the associated Green's function is not oscillatory. This work is in progress, and if successful will be reported at a later date.

References

[1] Ming Gu and Stanley C. Eisenstat, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput. **17** (1996), no. 4, 848–869.

[2] B. Alpert, G. Beylkin, R. Coifman, V. Rokhlin, *Wavelet-like bases for the fast solution of second-kind integral equations*, SIAM J. Sci. Comput., vol. 14, pp. 159-184, 1993.

[3] G. Beylkin, R. Coifman, and V. Rokhlin, *Fast wavelet transforms and numerical algorithms I*, Communications on Pure and Applied Mathematics, 14:141-183 (1991).

[4] Yu Chen, *Fast direct solver for the Lippmann-Schwinger equation*, Advances in Computational Mathematics, vol. 16, pp. 175-190, 2002.

[5] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[6] Å Björck, *Numerics of Gram-Schmidt orthogonalization*, Linear Algebra Appl., vol. 197/198, pp. 297–316, 1994.

[7] G.Beylkin, *On multiresolution methods in numerical analysis*, Documenta Mathematica, Extra Volume ICM 1998, III, pp. 481-490, 1998.

[8] G.W. Stewart, *Matrix Algorithms, Vol. I*, SIAM, Philadelphia 1998.

ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Computational Physics xxx (2004) xxx–xxx

JOURNAL OF COMPUTATIONAL PHYSICS

www.elsevier.com/locate/jcp

A fast direct solver for boundary integral equations in two dimensions

P.G. Martinsson *, V. Rokhlin

Department of Mathematics, Yale University, 10 Hillhouse Avenue, New Haven, CT 06511, USA

Received 7 May 2004; accepted 5 October 2004

Abstract

We describe an algorithm for the direct solution of systems of linear algebraic equations associated with the discretization of boundary integral equations with non-oscillatory kernels in two dimensions. The algorithm is "fast" in the sense that its asymptotic complexity is $O(n \log^{\kappa} n)$, where n is the number of nodes in the discretization, and κ depends on the kernel and the geometry of the contour ($\kappa = 1$ or 2). Unlike previous fast techniques based on iterative solvers, the present algorithm directly constructs a compressed factorization of the inverse of the matrix; thus it is suitable for problems involving relatively ill-conditioned matrices, and is particularly efficient in situations involving multiple right hand sides. The performance of the scheme is illustrated with several numerical examples.

1. Introduction

The boundary value problems of classical potential theory are ubiquitous in engineering and physics. Most such problems can be reduced to boundary integral equations which are, from a mathematical point of view, more tractable than the original differential equations. Although the mathematical benefits of such reformulations were realized and exploited in the 19th century, until recently boundary integral equations were rarely used as numerical tools, since most integral operators upon discretization turn into dense matrices. In the 1980s, the cost of applying dense matrices resulting from potential theory to arbitrary vectors was greatly reduced by the development of "fast" algorithms (Fast Multipole Methods, panel clustering, wavelets, etc.). Combining fast matrix-vector multiplication techniques with iterative schemes for the solution of large-scale systems of linear algebraic equations, it became possible to solve well-conditioned

* Corresponding author. Tel.: +1 203 432 1277.
E-mail address: per-gunnar.martinsson@yale.edu (P.G. Martinsson).

0021-9991/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2004.10.033

boundary integral equations of potential theory in $O(n)$ operations, where n is the number of unknowns. Today, such combinations are in many environments the fastest and most accurate numerical solution techniques available. Iterative linear solvers have certain drawbacks though; we briefly discuss these below.

(1) The number of iterations required by an iterative solver is sensitive to the spectral properties of the matrix of the system to be solved; for sufficiently ill-conditioned problems, the number of iterations is proportional to n. Since each iteration (with FMM acceleration) requires $O(n)$ operations, the total operation count is then proportional to n^2. While this is still better than the $O(n^3)$ estimate associated with direct solvers, in many situations $O(n^2)$ is not acceptable.
(2) When one needs to solve a collection of problems involving a single matrix and multiple right-hand sides, the CPU time requirements of most iterative algorithms are close to the time required to solve one problem multiplied by the number of problems to be solved. With most direct solvers, the situation is different; once the matrix has been inverted (or factored), applying its inverse to each additional right-hand side is very inexpensive.
(3) When a collection of linear systems has to be solved whose matrices are in some sense "close" to each other, iterative algorithms derive little (if any) advantage from the closeness of the matrices.
(4) Most direct schemes for the solution of linear systems are closely related to efficient algorithms for the construction of their Singular Value Decompositions and certain other matrix factorizations (L-U, Q-R, etc.). The simplest such scheme is probably the inverse power method with shifts (see, for example, [6]), which converts *any* algorithm for the solution of a linear system into an algorithm for the determination of a prescribed singular value. Iterative techniques do not provide such a capability, except via the so-called Lanczos schemes, which tend to require a large number of iterations (see, for example, [14]).

The subject of this paper is a numerical technique that is intended to overcome these shortcomings by directly producing a compressed ("data-sparse") factorization of the inverse of the matrix. When applied to contour integral equations of potential theory whose kernels are non-oscillatory, the asymptotic complexity of the solver is typically $O(n \log^{\kappa} n)$, where κ depends on the geometry and the kernel ($\kappa = 1$ or 2). When applied to problems involving oscillatory kernels, the asymptotic complexity deteriorates as the wavenumber increases but the scheme remains viable for objects up to a few hundred wavelengths in size. The factorization technique described in this paper is a multilevel extension of the compression technique described in [3]. The machinery underlying these techniques applies generally to matrices with rank-deficient off-diagonal submatrices; contour integral equations have been chosen by the authors simply as the most straightforward application.

It is not the purpose of this paper to provide an exhaustive survey of the literature on the subject we are addressing. A number of researchers have observed that matrices with rank-deficient off-diagonal blocks admit "fast" factorizations (see [8,9]); others have constructed "fast" algorithms in various environments (see [1,2,4,5,12]) where the operators in question posses rank-deficient off-diagonal blocks, without using this property explicitly. However, we observe that the algorithm of this paper is closely related to the scheme described in [13]. In fact, our algorithm could be viewed as a modification of the algorithm of [13] that replaces "elongated" objects in two or three dimensions with "curves", extends the class of kernels addressed by [13], and introduces modifications in the scheme of [13] that are necessary for this extension to work.

The paper is organized as follows: In Section 2 we introduce our notation and list certain facts about compression of rank-deficient matrices. In Section 3 we demonstrate that the inverse of a matrix with rank-deficient off-diagonal blocks possesses a data-sparse hierarchical factorization. In Section 4 we present a generic numerical technique for constructing the factorization described in Section 3. In Section 5 we show how the generic numerical technique presented in Section 4 can be improved further when applied

to contour integral equations. In Section 6 we illustrate through numerical examples the efficiency of the technique presented in Section 5 when applied to a number of different kernels and contours. In Section 7 we summarize our findings and discuss possible extensions and generalizations.

2. Preliminaries

2.1. Notation

Throughout the paper, we use upper case letters for matrices and lower case letters for vectors and scalars. The canonical unit vectors in $\mathbb{C}^n$ are denoted by e_j. Given a matrix $X \in \mathbb{C}^{m\times n}$, we let

X^* denote its adjoint (the complex conjugate transpose),
$\sigma_k(X)$ denote its kth singular value,
$\|X\|_2$ denote its l^2 operator norm,
$\|X\|_F$ denote its Frobenius norm, and
$x_j \in \mathbb{C}^{m\times 1}$ denote its jth column.

Given matrices A, B, C and D we let

$$[AB], \quad \begin{bmatrix} A \\ C \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{2.1}$$

denote larger matrices obtained by stringing the blocks A, B, C and D together.

Definition 1. (Permutation vectors) Given a positive integer n, we define

$$\mathbb{J}_n = \text{the set of permutations of the integers } \{1,\ldots,n\}. \tag{2.2}$$

Given two integers k and n such that $1 \leqslant k \leqslant n$, we define

$$\mathbb{J}_n^k = \text{the set of subsets of size } k \text{ of elements of } \mathbb{J}_n. \tag{2.3}$$

In other words, if $J \in \mathbb{J}_n^k$, then J is a vector of integers

$$J = [j_1,\ldots,j_k], \tag{2.4}$$

where $1 \leqslant j_l \leqslant n$ and all j_l's are different.

Definition 2. (Submatrix) When we use the term "submatrix" we do not insist that the submatrix must form a contiguous block. To be precise, we say that $B \in \mathbb{C}^{k\times l}$ is a submatrix of $A \in \mathbb{C}^{m\times n}$, if there exist permutations $I = [i_1,\ldots,i_k] \in \mathbb{J}_m^k$ and $J = [j_1,\ldots,j_l] \in \mathbb{J}_n^l$ such that

$$b_{pq} = a_{i_p j_q}, \quad \text{for } p = 1,\ldots,k, \ q = 1,\ldots,l. \tag{2.5}$$

Definition 3. (Neutered rows and columns) Let A be a matrix consisting of $p \times p$ blocks,

$$A = \begin{bmatrix} A^{(1,1)} & \cdots & A^{(1,p)} \\ \vdots & & \vdots \\ A^{(p,1)} & \cdots & A^{(p,p)} \end{bmatrix}. \tag{2.6}$$

We refer to the submatrix formed by all blocks on the ith row except the diagonal one, i.e.

$$\left[A^{(i,1)}\cdots A^{(i,i-1)}\ A^{(i,i+1)}\cdots A^{(i,p)}\right], \tag{2.7}$$

as the ith neutered row of blocks. A neutered column of blocks is defined analogously.

2.2. Compression of matrices

In this section we state a theorem on matrix compression that forms the foundation of the matrix factorization technique presented later in this paper. Roughly speaking, the theorem asserts that given a matrix A of rank k, it is possible to pick k of its columns in such a fashion that they form a well-conditioned basis for the remaining columns. It was first reported in slightly different form in [7].

Theorem 1. *Given an arbitrary matrix $A \in \mathbb{C}^{m\times n}$ and an integer k such that $1 \leqslant k < \min(m,n)$, there exists a (not necessarily unique) matrix $T \in \mathbb{C}^{k\times(n-k)}$ and a permutation $J = [j_1, \ldots, j_n] \in \mathbb{J}_n$ such that*

$$\tilde{A}_2 = \tilde{A}_1 T + E. \tag{2.8}$$

Here, $\tilde{A}_1$ and $\tilde{A}_2$ are matrices formed by the columns of A,

$$\begin{aligned} \tilde{A}_1 &= [a_{j_1}, \ldots, a_{j_k}] \in \mathbb{C}^{m\times k}, \\ \tilde{A}_2 &= [a_{j_{k+1}}, \ldots, a_{j_n}] \in \mathbb{C}^{m\times(n-k)}, \end{aligned} \tag{2.9}$$

the elements of the matrix $T \in \mathbb{C}^{k\times(n-k)}$ satisfy

$$|T_{ij}| \leqslant 1, \quad \text{for } 1 \leqslant i \leqslant k, \ 1 \leqslant j \leqslant n-k, \tag{2.10}$$

and the matrix $E \in \mathbb{C}^{m\times(n-k)}$ satisfies the inequality

$$\|E\|_2 \leqslant \sigma_{k+1}(A)\sqrt{1 + k(n-k)}, \tag{2.11}$$

where $\sigma_{k+1}(A)$ is the $(k+1)$th singular value of A.

Remark 4. (Computational complexity) While Theorem 1 asserts the theoretical existence of a matrix T and a permutation J with certain properties, it does not address the question of how to determine these numerically. In fact, the authors are not aware of any algorithm that finds these objects in polynomial time. However, in [7] an algorithm is presented that finds a matrix T and a permutation J such that all statements of Theorem 1 still hold, except that (2.10) and (2.11) are replaced by the weaker inequalities

$$|\, T_{ij} \,| \leqslant \sqrt{n}, \quad \text{for } 1 \leqslant i \leqslant k, \ 1 \leqslant j \leqslant n-k \tag{2.12}$$

and

$$\|E\|_2 \leqslant \sigma_{k+1}(A)\sqrt{1 + nk(n-k)}. \tag{2.13}$$

When $m \geqslant n$, the computational complexity of this algorithm is typically $O(mnk)$, the same as for the pivoted QR-factorization. In rare cases, the computational complexity may be somewhat larger but it never exceeds $O(mn^2)$.

Observation 5. (Column compression) When applied to a matrix $A \in \mathbb{C}^{m\times n}$ of rank k, Theorem 1 asserts that there exists a well-conditioned column operation that leaves k of the columns of A unchanged while mapping the remaining $n-k$ columns to zero. More specifically, let us define

$$R = P_J \begin{bmatrix} I & -T \\ 0 & I \end{bmatrix} \in \mathbb{C}^{n\times n}, \tag{2.14}$$

where T and J are defined in Theorem 1 and the permutation matrix P_J is defined by

$$P_J = [e_{j_1}, \ldots, e_{j_n}] \in \mathbb{C}^{n\times n}. \tag{2.15}$$

Then

$$AR = [A_{CS}0] \in \mathbb{C}^{m\times n}, \tag{2.16}$$

where the "column skeleton" A_{CS}, is formed by k of the columns of A;

$$A_{CS} = [a_{j_1}, \ldots, a_{j_k}] \in \mathbb{C}^{m\times k}. \tag{2.17}$$

Moreover, by virtue of (2.10) and the identity

$$R^{-1} = \begin{bmatrix} I & T \\ 0 & I \end{bmatrix} P_J^*, \tag{2.18}$$

it is clear that

$$\|R\|_F \leqslant \sqrt{n + k(n-k)}, \quad \text{and} \quad \|R^{-1}\|_F \leqslant \sqrt{n + k(n-k)}. \tag{2.19}$$

Observation 6. (Row compression) The argument of Observation 5 can equally well be applied to the rows of a matrix $A \in \mathbb{C}^{m\times n}$ of rank k. Thus, there exists a matrix $L \in \mathbb{R}^{m\times m}$ such that

$$LA = \begin{bmatrix} A_{RS} \\ 0 \end{bmatrix} \in \mathbb{C}^{m\times n}, \tag{2.20}$$

where the "row skeleton" $A_{RS} \in \mathbb{C}^{k\times n}$ is formed by k of the rows of A and

$$\|L\|_F \leqslant \sqrt{m + k(m-k)} \quad \text{and} \quad \|L^{-1}\|_F \leqslant \sqrt{m + k(m-k)}. \tag{2.21}$$

3. Analytical apparatus

Consider a $p \times p$ block matrix

$$A = \begin{bmatrix} A^{(11)} & \cdots & A^{(1p)} \\ \vdots & & \vdots \\ A^{(p1)} & \cdots & A^{(pp)} \end{bmatrix}, \tag{3.1}$$

such that any neutered row or column of blocks is rank-deficient. In this section we derive compressed factorizations of the inverse of such a matrix. Lemmas 2 and 3 provide factorizations for the case $p = 2$. Observation 8 extends the results of Lemma 3 to a general p. Observation 9 introduces hierarchical factorizations that further improve the efficiency.

Lemma 2 below asserts that for a given 2×2 block matrix with rank-deficient off-diagonal blocks, there exist well-conditioned row- and column-operations that (i) introduce zeros in the off-diagonal blocks and (ii) leave the remaining elements in the off-diagonal blocks untouched.

Lemma 2. *Let A be a non-singular matrix*

$$A = \begin{bmatrix} A^{(11)} & A^{(12)} \\ A^{(21)} & A^{(22)} \end{bmatrix} \in \mathbb{C}^{(n+m)\times(n+m)}, \tag{3.2}$$

where $A^{(11)} \in \mathbb{C}^{n\times n}, A^{(22)} \in \mathbb{C}^{m\times m}$ and the off-diagonal blocks $A^{(12)} \in \mathbb{C}^{n\times m}, A^{(21)} \in \mathbb{C}^{m\times n}$ have rank $k < \min(m,n)$. Then there exist matrices $R, L \in \mathbb{C}^{n\times n}$ such that

$$\begin{bmatrix} L & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{(11)} & A^{(12)} \\ A^{(21)} & A^{(22)} \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & A^{(12)}_{\mathrm{RS}} \\ X_{21} & X_{22} & 0 \\ A^{(21)}_{\mathrm{CS}} & 0 & A^{(22)} \end{bmatrix}. \tag{3.3}$$

Here, the matrix $A^{(12)}_{\mathrm{RS}} \in \mathbb{C}^{k\times m}$ consists of k of the rows of $A^{(12)}$ and the matrix $A^{(21)}_{\mathrm{CS}} \in \mathbb{C}^{m\times k}$ consists of k of the columns of $A^{(21)}$. Moreover, $X_{11} \in \mathbb{C}^{k\times k}$, $X_{12} \in \mathbb{R}^{k\times(n-k)}$, $X_{21} \in \mathbb{R}^{(n-k)\times k}$, $X_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$, and the matrices R and L satisfy (2.19) *and* (2.21), *respectively.*

Proof. Due to Observations 5 and 6, there exist matrices $R, L \in \mathbb{C}^{n\times n}$ such that

$$LA^{(12)} = \begin{bmatrix} A^{(12)}_{\mathrm{RS}} \\ 0 \end{bmatrix} \quad \text{and} \quad A^{(21)}R = \begin{bmatrix} A^{(21)}_{\mathrm{CS}} & 0 \end{bmatrix}, \tag{3.4}$$

where $A^{(12)}_{\mathrm{RS}}$ and $A^{(21)}_{\mathrm{CS}}$ are submatrices of $A^{(12)}$ and $A^{(21)}$, respectively. The identity (3.3) now follows by partitioning

$$\begin{aligned} &L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}, \quad \text{where } L_1 \in \mathbb{C}^{k\times n},\ L_2 \in \mathbb{C}^{(n-k)\times n}, \\ &R = [R_1\ R_2], \quad \text{where } R_1 \in \mathbb{C}^{n\times k},\ R_2 \in \mathbb{C}^{n\times(n-k)}, \end{aligned} \tag{3.5}$$

and setting

$$\begin{aligned} X_{11} &= L_1 A^{(11)} R_1 \in \mathbb{C}^{k\times k}, \\ X_{12} &= L_1 A^{(11)} R_2 \in \mathbb{C}^{k\times(n-k)}, \\ X_{21} &= L_2 A^{(11)} R_1 \in \mathbb{C}^{(n-k)\times k}, \\ X_{22} &= L_2 A^{(11)} R_2 \in \mathbb{C}^{(n-k)\times(n-k)}. \end{aligned} \tag{3.6}$$

□

The following lemma uses the results of Lemma 3 to reduce the problem of factoring the inverse of the matrix A in (3.2) to the problem of factoring the inverse of the smaller matrix $\tilde{A}$ in (3.8).

Lemma 3. *Let A, X_{11}, X_{12}, X_{21}, X_{22}, $A^{(12)}_{\mathrm{RS}}$ and $A^{(21)}_{\mathrm{CS}}$ be as in* Lemma 2. *Provided that the matrix X_{22} in* (3.3) *is non-singular, there exist matrices $B \in \mathbb{C}^{n\times k}$, $C \in \mathbb{C}^{k\times n}$ and $D \in \mathbb{C}^{n\times n}$ such that*

$$A^{-1} = \begin{bmatrix} B & 0 \\ 0 & I \end{bmatrix} \tilde{A}^{-1} \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.7}$$

where

$$\tilde{A} = \begin{bmatrix} \tilde{A}^{(11)} & A^{(12)}_{\mathrm{RS}} \\ A^{(21)}_{\mathrm{CS}} & A^{(22)} \end{bmatrix} \in \mathbb{C}^{(k+m)\times(k+m)} \tag{3.8}$$

and

$$\tilde{A}^{(11)} = X_{11} - X_{12}X_{22}^{-1}X_{21} \in \mathbb{C}^{k\times k}. \tag{3.9}$$

Proof. We let L_1, L_2, R_1 and R_2 be defined by (3.5). Inverting both sides of Eq. (3.3), we obtain the identity

$$A^{-1} = \begin{bmatrix} R_1 & R_2 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} & A^{(12)}_{\mathrm{RS}} \\ X_{21} & X_{22} & 0 \\ A^{(21)}_{\mathrm{CS}} & 0 & A^{(22)} \end{bmatrix}^{-1} \begin{bmatrix} L_1 & 0 \\ L_2 & 0 \\ 0 & I \end{bmatrix}. \tag{3.10}$$

Since X_{22} is non-singular,

$$\begin{bmatrix} X_{11} & X_{12} & A^{(12)}_{\mathrm{RS}} \\ X_{21} & X_{22} & 0 \\ A^{(21)}_{\mathrm{CS}} & 0 & A^{(22)} \end{bmatrix}^{-1} = \begin{bmatrix} I_k & 0 \\ -X_{22}^{-1}X_{21} & 0 \\ 0 & I_m \end{bmatrix} \begin{bmatrix} X_{11} - X_{12}X_{22}^{-1}X_{21} & A^{(12)}_{\mathrm{RS}} \\ A^{(21)}_{\mathrm{CS}} & A^{(22)} \end{bmatrix}^{-1} \times \begin{bmatrix} I_k & -X_{12}X_{22}^{-1} & 0 \\ 0 & 0 & I_m \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & X_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.11}$$

Now we obtain (3.7) by combining (3.10) and (3.11) and setting

$$\begin{aligned} B &= R_1 - R_2X_{22}^{-1}X_{21} \in \mathbb{C}^{n\times k}, \\ C &= L_1 - X_{12}X_{22}^{-1}L_2 \in \mathbb{C}^{k\times n}, \\ D &= R_2X_{22}^{-1}L_2 \in \mathbb{C}^{n\times n}. \quad \square \end{aligned} \tag{3.12}$$

Remark 7. (Symmetric factorizations) It is possible to force the factorization (3.7) to be symmetric in the sense that $R = L^*$ (which does not imply that $C = B^*$ unless A itself is Hermitian). To this end, we define L and J_{R} as the matrix and index vector that compress the rows of the matrix $[A^{(12)}A^{(21)^*}] \in \mathbb{R}^{n\times 2m}$ (rather than the rows of $A^{(12)}$ alone), and set $R = L^*$ and $J_{\mathrm{C}} = J_{\mathrm{R}}$. This modification typically results in a poorer compression ratio but may dramatically improve the conditioning of the transformation matrices, as discussed in Section 4.4.

Observation 8. (One-level compression of a block matrix) Consider a matrix

$$A = \begin{bmatrix} A^{(11)} & \cdots & A^{(1p)} \\ \vdots & & \vdots \\ A^{(p1)} & \cdots & A^{(pp)} \end{bmatrix} \in \mathbb{C}^{pn\times pn}, \tag{3.13}$$

where $A^{(ij)} \in \mathbb{C}^{n\times n}$ for $i,j = 1,\ldots,p$. We assume that any neutered row or column of blocks has rank at most k. Through p applications of Lemma 3, it is possible to reduce the problem of inverting A to the problem of inverting the smaller matrix

$$\tilde{A} = \begin{bmatrix} \tilde{A}^{(11)} & \cdots & \tilde{A}^{(1p)} \\ \vdots & & \vdots \\ \tilde{A}^{(p1)} & \cdots & \tilde{A}^{(pp)} \end{bmatrix} \in \mathbb{C}^{pk\times pk}, \tag{3.14}$$

where $\tilde{A}^{(ij)} \in \mathbb{C}^{k\times k}$ for $i,j = 1,\ldots,p$, and $\tilde{A}_{(ij)}$ is a submatrix of $A^{(ij)}$ whenever $i \neq j$.

More specifically, applying Lemma 3 to each of the p diagonal blocks of A, we obtain the factorization

$$A^{-1} = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & B_p \end{bmatrix} \tilde{A}^{-1} \begin{bmatrix} C_1 & 0 & \cdots & 0 \\ 0 & C_2 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & C_p \end{bmatrix} + \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & D_p \end{bmatrix}, \tag{3.15}$$

where $B_i \in \mathbb{C}^{n\times k}$, $C_i \in \mathbb{C}^{k\times n}$ and $D_i \in \mathbb{C}^{n\times n}$, for $i = 1, \ldots, p$.

The single-level matrix compression is illustrated graphically in Fig. 1.

Observation 9. (Hierarchical compression of a block matrix) Observation 8 reduces the problem of inversion of a block matrix with rank-deficient neutered rows and columns to the problem of inversion of a block matrix with smaller blocks. If by clustering these smaller blocks, we can create a matrix with off-diagonal rank-deficiencies, then the process can be repeated recursively to further improve the compression.

More specifically, let us change notation so that the objects labeled A, $\tilde{A}$ and k in Observation 8 are now labeled $A^{(1)}$, $\tilde{A}^{(1)}$ and k_1, respectively. Eq. (3.15) then reads

$$\left(A^{(1)}\right)^{-1} = B^{(1)}\left(\tilde{A}^{(1)}\right)^{-1} C^{(1)} + D^{(1)}, \tag{3.16}$$

where $B^{(1)}$, $C^{(1)}$, $D^{(1)}$ are block diagonal matrices whose p diagonal blocks are of sizes $n \times k_1$, $k_1 \times n$, $n \times n$, respectively. We then cluster the blocks of the matrix $\tilde{A}^{(1)}$ to form a matrix $A^{(2)}$ with $(p/2) \times (p/2)$ blocks of size $2k_1 \times 2k_1$ and apply the factorization (3.16) to it, thus obtaining a telescoped factorization

$$\left(A^{(1)}\right)^{-1} = B^{(1)}\left[B^{(2)}\left(\tilde{A}^{(2)}\right)^{-1} C^{(2)} + D^{(2)}\right] C^{(1)} + D^{(1)}. \tag{3.17}$$

Here, $A^{(2)}, B^{(2)}, C^{(2)}, D^{(2)}$ are all block matrices with $(p/2) \times (p/2)$ blocks. Letting k_2 denote the rank of the neutered rows and columns of $A^{(2)}$, the blocks of $\tilde{A}_{(2)}$ have size $k_2 \times k_2$, while $B^{(2)}$, $C^{(2)}$, $D^{(2)}$ are diagonal block matrices with diagonal blocks of sizes $2k_1 \times k_2$, $k_2 \times 2k_1$ and $2k_1 \times 2k_1$, respectively. This process can be continued until no further clustering is advantageous.

The multi-level matrix compression is illustrated graphically in Fig. 2.

Remark 10. (Adjoint of the inverse) Obviously, the factorizations (3.15) and (3.17) provide a mechanism for the accelerated application of both A^{-1} and $[A^{-1}]^*$.

Remark 11. (Block sizes) In Observations 8 and 9, it was assumed that all blocks within one of the matrices $A, \tilde{A}, A^{(1)}, A^{(2)}, \ldots$, have the same size. This assumption was made for notational convenience only and is in no way essential to the results.

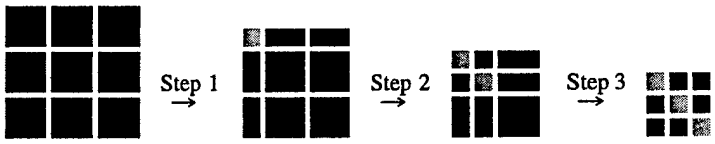


Fig. 1. A 3×3 matrix $[A^{(ij)}]_{i,j=1}^{3}$ is compressed in three steps, cf. Observation 8. In step $j = 1, 2, 3$, the single-block compression of Lemma 3 is applied to compress the interaction between $A^{(ij)}$ and the rest of the matrix. *Black* blocks represents entries that have not been changed beyond row and column permutations and *gray* represents entries that have been updated but are not (necessarily) zero.

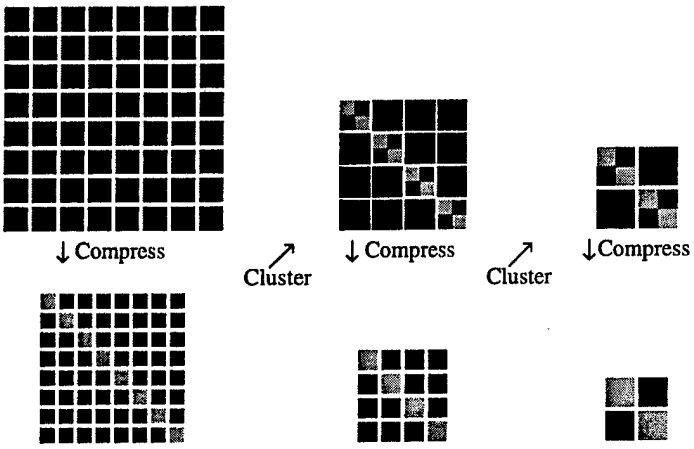


Fig. 2. An 8×8 block matrix is compressed through a three-level compression scheme in the vein of Observation 9. The gray scale coding is the same as in Fig. 1.

4. An algorithm for the computation of a compressed inverse

In Section 3 we demonstrate the existence of a compact factorization of the inverse of any block matrix whose neutered rows and columns of blocks are rank-deficient. In this section, we describe a numerical scheme for the construction of such factorizations, and estimate its efficiency.

Remark 12. The inversion scheme presented in this section is fairly generic, depending only on the ranks of off-diagonal blocks of the matrix to be inverted. In situations where the structure of the matrix is known, further improvements are possible. For instance, when applied to a dense $n \times n$ matrix resulting from the discretization of a contour integral operator, the generic algorithm of this section requires $O(n^2)$ arithmetic operations to construct its inverse, while the customized technique presented in Section 5 requires $O(n \log^2 n)$ operations or less, depending on the integral operator.

4.1. Single block compression

Lemmas 2 and 3 assert that the inverse of a 2×2 block matrix of the form (3.2) can be factored in the compressed form (3.7). The quantities $\tilde{A}_{(11)}$, R, L, $A^{(12)}_{\mathrm{RS}}$ and $A^{(21)}_{\mathrm{CS}}$ that appear in (3.7) can be determined by taking the following steps:

(1) Determine a matrix $L \in \mathbb{C}^{n \times n}$ and a permutation $J_{\mathrm{R}} \in \mathbb{J}^k_n$ such that

$$LA^{(12)} = \begin{bmatrix} A^{(12)}_{\mathrm{RS}} \\ 0 \end{bmatrix},$$

where $A^{(12)}_{\mathrm{RS}}$ is formed by the k rows of $A^{(12)}$ specified by J_{R}, as described in Observation 6.

(2) Determine a matrix $R \in \mathbb{C}^{n \times n}$ and a permutation $J_{\mathrm{C}} \in \mathbb{J}^k_n$ such that

$$A^{(21)}R = \begin{bmatrix} A^{(21)}_{\mathrm{CS}} & 0 \end{bmatrix},$$

where $A^{(21)}_{\mathrm{CS}}$ is formed by the columns of $A^{(21)}$ specified by J_{C}, as described in Observation 5.

(3) Partition R and L as specified in (3.5) and form the blocks X_{ij} as in (3.6).
(4) Compute $\tilde{A}_{(11)}$, B, C and D using the formulas (3.9) and (3.12).

Steps (1) and (2) require $O(mnk)$ floating point operations while steps (3) and (4) require $O(n^3)$ operations. The total cost is thus $O(mnk + n^3)$.

4.2. Single-level compression

Let A denote a matrix consisting of $p \times p$ blocks, each of size $n \times n$, in which every neutered row or column has rank k such that $k < n$. Observation 8 states that such a matrix can be factored in the sparse form (3.15). This factorization contains the entities B_i, C_i, D_i, $\tilde{A}_{(ij)}$ for $i,j = 1, \ldots, p$, which can be computed through p applications of the single-block compression technique of Section 4.1 – one application for each diagonal block. Each one of the p steps requires $O(pkn^2 + n^3)$ floating point operations resulting in a total computational cost of $O(p^2kn^2 + pn^3)$.

Remark 13. The off-diagonal blocks of the compressed matrix $\tilde{A}$ are never explicitly computed. Instead, the block $\tilde{A}^{(ij)} \in \mathbb{C}^{k\times k}$ is specified by giving the index vectors $J_R^{(i)}, J_C^{(j)} \in \mathbb{J}_n^k$ that define the rows and columns of $A^{(ij)} \in \mathbb{C}^{k\times k}$, whose intersections form $\tilde{A}^{(ij)}$. (Here $J_R^{(i)}$ is the index vector obtained when compressing the ith row of blocks and $J_C^{(j)}$ is the index vector obtained when compressing the jth column of blocks.)

4.3. Multi-level compression

The single-level technique compresses a block matrix A to form another block matrix $\tilde{A}$ with smaller blocks. Now, if by clustering blocks, we can create rank-deficiencies in the neutered rows and columns of $\tilde{A}$, then the single-level technique can be applied recursively. The algorithmic implementation entirely follows the description in Observation 9.

When estimating the computational cost for the multi-level technique we use $r = 1, \ldots, R$ as an index for the levels (with $r = 1$ being the finest level), we let p_r denote the number of blocks on level r, n_r the average block size and k_r the average rank. The cost for step r is then

$$t_r \sim k_r p_r^2 n_r^2 + p_r n_r^3. \tag{4.1}$$

We assume that $p_r k_r \geqslant n_r$ so that the second term is dominated by the first. Using that $p_r k_r = p_{r+1} n_{r+1}$, we then find that the total cost for all R steps is

$$T \sim \sum_{r=1}^{R} t_r \sim \sum_{r=1}^{R} p_{r+1} p_r n_{r+1} n_r^2. \tag{4.2}$$

At each level, the number of blocks is cut in half, so

$$p_r = \frac{p_1}{2^{r-1}}. \tag{4.3}$$

We let $\gamma_r = n_{r+1}/2n_r$ denote the compression ratio so that

$$n_r = (2\gamma_{r-1}) \cdots (2\gamma_1) n_1. \tag{4.4}$$

Assuming that there exists a constant γ such that $\gamma_r \leqslant \gamma$, we obtain the bound

$$n_r \leqslant (2\gamma)^{r-1} n_1. \tag{4.5}$$

Combining (4.2), (4.3) and (4.5), we find that the total cost is

$$T \sim \sum_{r=1}^{R} \frac{p_1}{2^r} \frac{p_1}{2^{r-1}} (2\gamma)^r n_1 (2\gamma)^{2r-2} n_1^2 \sim p_1^2 n_1^3 \sum_{r=1}^{R} (2\gamma^3)^r. \tag{4.6}$$

We assume that $\gamma < \sqrt[3]{4} = 0.7937\cdots$ so that the sum is bounded by $(1 - 2\gamma^3)^{-1}$. Letting N denote the size of the matrix we find that $N = p_1 n_1$ and thus

$$T \sim N^2 n_1. \tag{4.7}$$

The assumption that (4.5) holds for some $\gamma < 0.7939\cdots$ is valid in many environments relating to discretization of contour integral equations. We will return to this point in Section 6.

4.4. Conditioning

All factorizations computed in this section are variations of (3.15). For this formula to be of practical use, the matrices B_i, C_i and D_i must not be excessively large (in say the l^2 operator norm) and the condition number of $\tilde{A}$ has to be similar to that of A. The formulas (3.12) imply that this is true if $\|X_{22}^{-1}\|_2$ is of moderate size (since (2.19) and (2.21) assert that R and L are well-conditioned). Under the assumptions of this section (that the global matrix be non-singular and the off-diagonal blocks have low rank) it is not possible to prove any such bound.

However, in the context of contour integral equations, the problem can largely be avoided by enforcing that the compression be symmetric in the sense of Remark 7. The reason is that the diagonal blocks of the original matrix tend to have the form

$$A^{(11)} = D + E, \tag{4.8}$$

where D is a positive definite Hermitian matrix and E is "small" compared to D in operator norm. Since $R_2 = L_2^*$ when symmetry is enforced, we find that, cf. (3.6),

$$X_{22} = L_2(D+E)L_2^* = (L_2 D^{1/2})(L_2 D^{1/2})^* + L_2 E L_2^*. \tag{4.9}$$

Here, the first term is well-conditioned, and the second has at most a few non-small singular values. Thus, it is very unlikely that the sum of the two matrices should have any small singular values. Furthermore, should such a coincidence happen, the algorithm detects it and avoids the problem by locally re-partitioning the matrix.

4.5. Error estimation

Given a prescribed accuracy ε, the numerical scheme presented in this section solves the equation

$$Au = f \tag{4.10}$$

by constructing an approximation A_ε that satisfies

$$\|A - A_\varepsilon\|_2 \leqslant \varepsilon \tag{4.11}$$

and is such that the approximate solution $u_\varepsilon = A_\varepsilon^{-1} f$ can be computed fast. The error in u satisfies

$$u - u_\varepsilon = (A^{-1} - A_\varepsilon^{-1}) f = A_\varepsilon^{-1}(A_\varepsilon - A)A^{-1} f = A_\varepsilon^{-1}(A_\varepsilon - A)u. \tag{4.12}$$

The relative error is therefore bounded as follows:

$$\frac{\|u - u_\varepsilon\|}{\|u\|} \leqslant \|A_\varepsilon^{-1}(A_\varepsilon - A)\|_2 \leqslant \varepsilon \|A_\varepsilon^{-1}\|_2. \tag{4.13}$$

While the algorithm cannot possibly control $\|A_\varepsilon^{-1}\|_2$, this quantity can be computed cheaply using power iteration, see Remark 10. Thus, an assured bound for the relative error can be computed à posteriori.

5. An accelerated algorithm applicable to contour integral equations

The bulk of the computational cost of the matrix compression technique presented in Section 4 consists of the cost of determining index vectors and transformation matrices that compress the neutered rows and columns. When the matrix under consideration is a discrete approximation of a contour integral operator, it is possible to determine these quantities through an entirely local operation whose cost only depends on the size of the diagonal block to be compressed (i.e., not on the size of the rest of the matrix). This is possible since the column and row operations employed in the present matrix compression technique do not update the elements of the off-diagonal blocks, as discussed in Remark 13.

This section is structured as follows: In Section 5.1 we describe a single-block compression technique for the boundary integral equations associated with Laplace's equation in two dimensions that is faster than the generic single-block technique of Section 4.1. In Section 5.2 we describe single and multi-level techniques for contour integral equations obtained by repeated application of the single-block compression technique of Section 5.1. Section 5.3 discusses generalizations of the technique to other equations of potential theory.

Remark 14. (Numerically rank-deficient matrices) In this section, we say that a matrix has rank k provided that it has only k singular values that are larger than some preset accuracy. In other words, we do not distinguish between what is sometimes called "numerical rank" and actual rank.

5.1. Single-block compression

The following observation summarizes the principle finding of this section:

Observation 15. Let the matrix A in (3.2) represent the discretization of the integral operator

$$\int_\Gamma K(x,y)u(y)\mathrm{d}s(y), \quad \text{for } x \in \Gamma, \tag{5.1}$$

where $\Gamma = \Gamma_1 + \Gamma_2$ is a contour (Fig. 3 shows one example), the block structure of A corresponds to the partitioning of Γ (so that, e.g., $A^{(12)}$ represents evaluation on Γ_1 of the potential generated by a charge distribution on Γ_2), and K is the kernel of a single and/or double layer potential for the Laplace operator. Then under very mild assumptions on the contour Γ, the factorization (3.3) can be computed using $O(n^3)$ floating point operations, where n is the number of points used in the discretization of Γ_1.

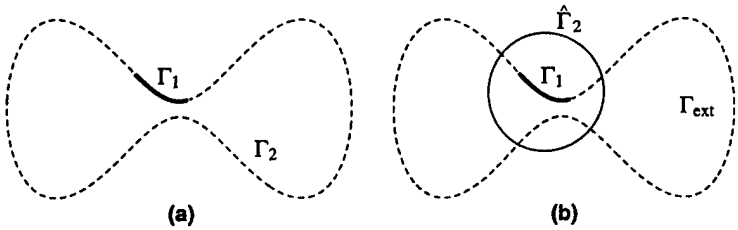


Fig. 3. A contour Γ. In figure (a), the partitioning $\Gamma = \Gamma_1 + \Gamma_2$ is shown with Γ_1 drawn with a bold line. In figure (b) the contour $\hat{\Gamma}_2$ is drawn with a thin solid line and Γ_{ext} with a dashed line.

The idea behind the construction alluded to in Observation 15 is simple: Instead of compressing the interaction between Γ_1 and Γ_2, it is sufficient to compress the interaction between Γ_1 and a small contour $\hat{\Gamma}_2$, formed by the union of an artificial circular contour enclosing Γ_1 and the part of Γ_2 that is inside this circle (as shown in Fig. 3(b)). The reason is that by virtue of Green's theorem, any potential field generated by charges on Γ_2 can equally well be generated by charges on $\hat{\Gamma}_2$. Finally we note that if Γ_1 is discretized using n nodes, then typically $\hat{\Gamma}_2$ can be discretized using $O(n)$ nodes, yielding a total cost for the procedure of $O(n^3)$.

The remainder of this subsection is devoted to substantiating Observation 15. We start by introducing some notation; let Γ_{circ} denote the circle in Fig. 3(b) and let Γ_{ext} denote the part of Γ_2 outside of Γ_{circ}. Furthermore, let $\boldsymbol{S}_{\Gamma_2\to\Gamma_1}$ denote the integral operator that evaluates a potential on Γ_1 caused by a charge distribution on Γ_2. In other words, $\boldsymbol{S}_{\Gamma_2\to\Gamma_1}$ acts on a charge distribution u as follows:

$$[\mathbf{S}_{\Gamma_2\to\Gamma_1}u](x) = \int_{\Gamma_2} K(x,y)u(y)\mathrm{d}s(y), \quad \text{for } x \in \Gamma_1. \tag{5.2}$$

Observation 15 rests on the following claim:

Lemma 4. *Let $H \in \mathbb{C}^{n\times n'}$ denote the matrix discretizing $\boldsymbol{S}_{\Gamma_{\text{circ}}\to\Gamma_1}$, and let the index vector $J_{\text{R}} \in \mathbb{J}_n^k$ and the transformation matrix L be such that they compress H in the sense of* Observation 6. *Then J_{R} and L also compress the matrix $B \in \mathbb{C}^{n\times m}$ that approximates the operator $\boldsymbol{S}_{\Gamma_{\text{ext}}\to\Gamma_1}$.*

Sketch of proof. It is sufficient to prove that there exists a matrix $W \in \mathbb{C}^{n'\times m}$ with moderate l^2 operator norm such that

$$B = HW. \tag{5.3}$$

(The matrix W is the matrix that maps a charge distribution on Γ_{ext} to an equivalent charge distribution on Γ_{circ}.) Now, Eq. (5.3) is the discrete approximation of the operator relation

$$\mathbf{S}_{\Gamma_{\text{ext}}\to\Gamma_1} = \mathbf{S}_{\Gamma_{\text{circ}}\to\Gamma_1}\left[(\mathbf{S}_{\Gamma_{\text{circ}}\to\Gamma_{\text{circ}}})^{-1}\mathbf{S}_{\Gamma_{\text{ext}}\to\Gamma_{\text{circ}}}\right]. \tag{5.4}$$

The matrix W in (5.3) corresponds to the operator in square brackets in (5.4). That this operator is bounded is a consequence of Green's theorem. □

5.2. Single- and multi-level compression

The generic single- and multi-level compression techniques of Sections 4.2 and 4.3 were obtained by repeated application of the single-block technique described in Section 4.1. Single- and multi-level techniques for contour integral equations are analogously obtained by repeated application of the single-block technique of Section 5.1.

It remains to estimate the computational cost of the accelerated compression technique. The cost for a single level compression at level $r = 1, \ldots, R$ is now, cf. (4.1),

$$t_r \sim p_r n_r^3, \tag{5.5}$$

where p_r denotes the number of clusters on level r and n_r is the (average) cluster size. Under the assumptions (4.3) and (4.5), we find that

$$t_r \sim \frac{p_1}{2^{r-1}}(2\gamma)^{3r-3}n_1^3. \tag{5.6}$$

The total cost for all R steps is then

$$T \sim \sum_{r=1}^{R} p_r n_r^3 \leqslant p_1 n_1^3 \sum_{r=1}^{R} \left(4\gamma^3\right)^{r-1}. \tag{5.7}$$

We assume that $\gamma < 4^{-1/3} = 0.630\cdots$ so that the sum is bounded by $(1 - 4\gamma^3)^{-1}$. Letting N denote the size of the original matrix, we find that $N = n_1 p_1$ and thus

$$T \sim N n_1^2. \tag{5.8}$$

When the kernel of the equation is associated with the fundamental solution of Laplace's equation, it is possible to prove that the assumption (4.5) holds with $\gamma \approx 1/2$ when $n_1 \geqslant \log N$, which gives an upper bound on the computational cost of $O(N \log^2 N)$. However, further acceleration is achieved by choosing a smaller n_1, even though the cluster size then grows slightly in the first couple of compressions. This explains why the $\log^2 N$ factor is not visible in the experiments in Section 6.

Remark 16. The single-block compression technique described in Observation 15 requires the algorithm to determine which of the nodes of Γ_2 lie inside the artificial circle Γ_{circ}. If this search would be done by brute force, the computational cost for a single level solve would include a term $p_r^2 n_r^2$, cf. (5.5). Even though the constant in front of this term is small, it would dominate the computation for large problems (in our implementation, this would happen for $N \geqslant 25000$). One solution to this problem is to perform the search via a hierarchical search tree; the estimate (5.5) then remains valid.

5.3. Generalizations

The technique presented in Section 5.1 for Laplace's equation is readily applicable to other equations of classical potential theory; Helmholtz, Yukawa, Shrödinger, Maxwell, Stokes, elasticity, etc. The only complication occurs when working with equations that may have resonances. In such cases, it is possible that the operator of self-interaction for the artificial circle (the operator $S_{\Gamma_{\text{circ}} \to \Gamma_{\text{circ}}}$ in (5.4)) has a non-trivial null-space. This complication can be rectified by letting the artificial charges on Γ_{circ} consist of both monopoles and dipoles. Alternatively, it is possible to consider only one type of charges but placing them on two concentric circles instead of a single one.

When applied to oscillatory problems such as Helmholtz' and Maxwell's equations, the efficiency of the technique deteriorates when the wave number increases since then the compression rate deteriorates as the blocks grow larger (in other words, the assumption (4.5) no longer holds). In practice, it appears that the method experiences very few problems for objects smaller than about 50 wavelengths. After that, the computational complexity increases superlinearly with the problem size although the technique remains viable for equations set on contours a few hundred wavelengths in size. This effect will be illustrated in the numerical examples in Section 6.1.

Finally we remark that the scheme has $O(n \log^\kappa n)$ complexity when applied to integral equations defined on one-dimensional curves in *any* dimension. The fact that we have so far only discussed equations embedded in two space dimensions is simply that contour integral equations associated with boundary value problems in two dimensions is the most common source of such equations.

6. Numerical examples

In this section we present the results of a number of numerical experiments performed to assess the efficiency of the numerical scheme presented in Sections 4 and 5. In every experiment, we compute a compressed factorization of the inverse of the matrix resulting from Nyström discretization of one of the following three integral equations:

$$\pm\frac{1}{2}u(x)+\frac{1}{2\pi}\int_\Gamma [n(y)\cdot\nabla_y \log|x-y|]u(y)\mathrm{d}s(y)=f(x),\quad x\in\Gamma, \tag{6.1}$$

$$\int_\Gamma [\log|x-y|]u(y)\mathrm{d}s(y)=f(x),\quad x\in\Gamma, \tag{6.2}$$

$$\mp 2\mathrm{i}u(x)+\int_\Gamma [(n(y)\cdot\nabla_y+\mathrm{i}k)H_0(k|x-y|)]u(y)\mathrm{d}s(y)=f(x),\quad x\in\Gamma, \tag{6.3}$$

where $n(y)$ is the outward pointing unit normal of Γ at y and $H_0(x)=J_0(x)+\mathrm{i}Y_0(x)$ is the Hankel function of zeroth order. Eqs. (6.1) and (6.2) are the double and single layer equations associated with Laplace Dirichlet problems, and (6.3) is an equation associated with the Helmholtz Dirichlet problem with wave number k. In Eqs. (6.1) and (6.3), the top sign in front of the first term refers to exterior problems and the lower sign refers to interior problems.

The kernel in (6.1) is smooth and the equation was discretized using the trapezoidal rule (which is exponentially convergent on a smooth contour). The Eqs. (6.2) and (6.3) involve log-singular kernels that were discretized using the modified trapezoidal quadrature rules of [10] of orders 6 and 10, respectively. The algorithm was implemented in Fortran 77 and the experiments were run on a 2.8 GHz Pentium 4 desktop with 512Mb of RAM memory.

When presenting the numerical results, we use the following notation:

R	the number of levels in the multi-level solver,
N_{start}	the size of the discrete problem at the start,
N_{final}	the size of the compressed problem,
t_{tot}	the total CPU time (in seconds),
t_{solve}	the CPU time required to apply the factorized inverse (in seconds),
c_{top}	the condition number of the compressed matrix,
σ_{min}	the smallest singular value of the original matrix,
M	the amount of memory used (in MB),
E_{actual}	the relative error in u, $E_{\text{actual}}=\|u_\varepsilon-u\|/\|u\|$,
E_{res}	the relative residual error, $E_{\text{res}}=\|Au_\varepsilon-f\|/\|f\|$,

In each experiment, the right hand side f was the Dirichlet data corresponding to a potential field generated by a few randomly placed point charges. Since the exact potential field was known, we could compare the potential field generated by the numerical solution to the exact one. We did this at J random points on a circle enclosing Γ and separated from Γ by one quarter of its radius. Letting $\{v^{(j)}\}_{j=1}^J$ denote the exact potential and $\{v_\varepsilon^{(j)}\}_{j=1}^J$ denote the potential generated by u_ε, we define the relative l^2-norm error in the potential as $E_{\text{pot}}=\|v-v_\varepsilon\|/\|v\|$.

6.1. Example: a smooth contour

In this subsection we present results pertaining to the smooth contour shown in Fig. 4. The contour was discretized using between 800 and 102 400 points and the integral equations associated with exterior Dirichlet problems were solved. Tables 1–3 present the results for the kernels (6.1)–(6.3), respectively. As a reference, we give in Table 4 the timings for highly optimized implementations of the LU-factorization, direct matrix-vector multiplication and FMM-accelerated matrix-vector multiplication.

For the two Laplace problems considered, we see that both the computational cost and the memory requirement scale more or less linearly with the problem size, as expected. We recall that this expectation was based on the postulate that for Laplace problems, the interaction rank between adjacent clusters de-

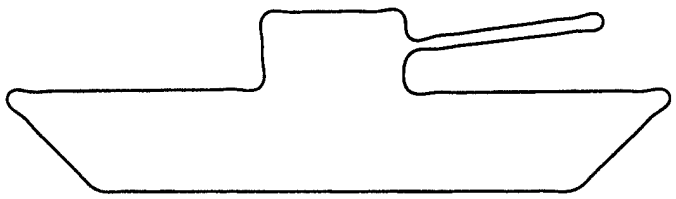


Fig. 4. A smooth contour. The length of the contour is roughly 5.1 and its horizontal width is 2.

Table 1
Computational results for the double layer potential (6.1) associated with an exterior Laplace Dirichlet problem on the contour shown in Fig. 4

N_{start}	N_{final}	t_{tot}	t_{solve}	E_{actual}	E_{res}	E_{pot}	c_{top}	σ_{min}	M
400	301	5.3e−01	2.9e−03	2.3e−10	4.7e−10	3.0e−06	4.3e+00	1.3e−02	4.2e+00
800	351	9.6e−01	4.1e−03	2.5e−10	2.2e−10	6.3e−10	9.1e+00	1.2e−02	6.5e+00
1600	391	1.6e+00	6.3e−03	1.4e−10	1.3e−10	1.6e−10	1.6e+01	1.2e−02	9.2e+00
3200	391	1.8e+00	8.5e−03	–	6.6e−11	3.7e−10	3.2e+01	1.2e−02	1.1e+01
6400	391	2.2e+00	1.2e−02	–	5.9e−11	8.9e−11	7.7e+01	1.2e−02	1.4e+01
12 800	390	2.6e+00	1.9e−02	–	3.6e−11	5.9e−11	1.4e+02	1.2e−02	2.1e+01
25 600	391	3.9e+00	3.4e−02	–	2.7e−11	4.7e−10	2.1e+02	–	3.5e+01
51 200	393	6.5e+00	6.5e−02	–	2.5e−11	5.3e−11	2.0e+02	–	6.3e+01
102 400	402	1.3e+01	1.2e−01	–	2.0e−11	–	1.1e+03	–	1.2e+02

Table 2
Computational results for the single layer potential (6.2) associated with an exterior Laplace Dirichlet problem on the contour shown in Fig. 4

N_{start}	N_{final}	t_{tot}	t_{solve}	E_{actual}	E_{res}	E_{pot}	c_{top}	σ_{min}	M
400	253	4.1e−01	1.9e−03	4.6e−09	2.7e−09	1.6e−04	2.2e+01	3.5e−02	3.1e+00
800	306	8.2e−01	3.3e−03	7.5e−09	9.9e−09	2.4e−06	1.6e+02	2.9e−04	5.4e+00
1600	353	1.6e+00	6.2e−03	4.9e−09	6.3e−09	1.6e−09	1.5e+02	1.4e−04	8.6e+00
3200	369	2.3e+00	9.7e−03	–	2.5e−07	1.2e−10	2.1e+02	4.2e−05	1.2e+01
6400	379	3.2e+00	1.6e−02	–	1.3e−08	6.8e−12	2.6e+02	2.1e−05	1.8e+01
12 800	395	4.8e+00	2.7e−02	–	1.7e−08	3.4e−12	2.8e+02	2.7e−06	2.9e+01
25 600	409	7.7e+00	4.8e−02	–	3.6e−08	1.4e−11	3.5e+02	2.7e−07	5.0e+01
51 200	419	1.4e+01	9.0e−02	–	2.7e−07	–	3.7e+02	3.5e−07	9.1e+01
102 400	429	3.6e+01	1.7e−01	–	1.6e−08	–	5.2e+02	–	1.7e+02

Table 3
Computational results for the kernel (6.3) associated with an exterior Helmholtz Dirichlet problem on the contour shown in Fig. 4

k	N_{start}	N_{final}	t_{tot}	t_{solve}	E_{actual}	E_{res}	E_{pot}	c_{top}	σ_{min}	M
21	800	435	1.5e+01	3.3e−02	2.7e−07	9.7e−08	7.1e−07	4.1e+03	6.5e−01	1.3e+01
40	1600	550	3.0e+01	6.7e−02	1.6e−07	6.2e−08	4.0e−08	6.1e+03	8.0e−01	2.5e+01
79	3200	683	5.3e+01	1.2e−01	–	5.3e−08	3.8e−08	2.1e+04	3.4e−01	4.5e+01
158	6400	870	9.2e+01	2.0e−01	–	3.9e−08	2.9e−08	4.0e+04	3.4e−01	8.2e+01
316	12 800	1179	1.8e+02	3.9e−01	–	2.3e−08	2.0e−08	4.2e+04	3.4e−01	1.6e+02
632	25 600	1753	4.3e+02	7.5e+00	–	1.7e−08	1.4e−08	9.0e+04	3.3e−01	3.5e+02
1264	51 200	2864	(1.5e+03)	(2.3e+02)	–	9.5e−09	–	–	–	8.3e+02

The Helmholtz parameter was chosen to keep the number of discretization points per wavelength constant at roughly 45 points per wavelength (resulting in a quadrature error about 10^{-12}). The times in parenthesis refer to experiments that did not fit in RAM.

Table 4
Timings (in seconds) for highly optimized implementations of the LU-factorization, matrix-vector multiplication and FMM accelerated matrix-vector multiplication

N	400	800	1600	3200	6400	12 800	25 600	51 200	102 400
t_{LU}	2.8e−02	2.0e−01	1.6e+00	1.3e+01	(1.0e+02)	(8.3e+02)	(6.7e+03)	(5.3e+04)	(4.3e+05)
t_{mult}	7.5e−04	2.9e−03	1.2e−02	4.8e−02	(1.9e−02)	(7.7e−01)	(3.1e+00)	(1.2e+01)	(4.9e+01)
t_{FMM}	3.8e−03	8.0e−03	1.6e−02	3.0e−02	6.0e−02	1.2e−01	2.4e−01	4.8e−01	9.6e−01

The FMM was run at a relative accuracy of 10^{-10} with the same kernel as in the Eq. (6.2). The numbers in parenthesis are extrapolated.

pend only very weakly (logarithmically) on their size. Fig. 5 illustrates this point; it shows that after two rounds of compression, almost the only nodes that have survived are the ones near the border to the neighboring clusters. The figure also illustrates that the algorithm detects the need to keep more nodes in the interior of those clusters that run close to other clusters. (For an example of a situation where the Eq. (6.1) needs to be discretized using a large number of nodes in spite of the fact that the contour is uncomplicated, see [11].)

Since the scheme presented in this paper relies on rank-considerations only, it works for oscillatory problems with low wave numbers but it eventually fails as the wavenumber is increased. Table 5 illustrates this point by showing how the compression ratios deteriorate as the wavenumber k in Eq. (6.3) is increased from 1 to 1200. However, the authors were surprised to find that the method remains viable up to objects about 200 wavelengths across, as indicated in Table 3.

Remark 17. (Comparison with the fast multipole method) From Tables 1 and 4, we see that a single FMM matrix-vector multiply is about 15–20 times faster than a matrix inversion. Thus, if an iterative solver requires less than 15–20 iterations to solve Eq. (6.1), this would beat the direct method for a single solve. However, once the inverse has been computed, it can be applied to additional right hand sides in about one tenth of the time required for a single FMM accelerated matrix-vector multiply.

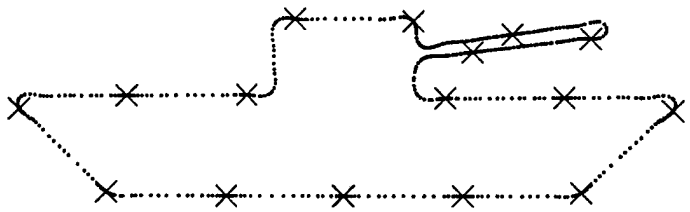


Fig. 5. The points left after two rounds of compression of the contour shown in Fig. 4. The crosses mark the boundary points between adjacent clusters.

Table 5
Deterioration of compression rates for large wavenumber Helmholtz problems

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	N_{final}	M
1	0.68	0.58	0.54	0.55	0.58	0.64	0.64	0.72	512	167
100	0.72	0.56	0.55	0.56	0.60	0.68	0.72	0.82	777	197
500	0.72	0.58	0.58	0.62	0.68	0.76	0.84	0.91	1522	303

The table shows the compression ratio γ_j, see (4.4), at each of the levels $j = 1, \ldots, 8$ for the Helmholtz kernel (6.3) on the smooth contour in Fig. 4, discretized with $N = 25\,600$ points. The three rows correspond to wave numbers $k = 1$, 100, 500. The second to last column shows the number of degrees of freedom left on the finest level and the last column shows the total memory requirement (in MB).

6.2. A rippled contour that almost self-intersects

In this subsection we present results pertaining to the rippled contour shown in Fig. 6. The contour was discretized using between 800 and 102 400 points and integral equations associated with exterior Dirichlet problems were solved. The number of ripples in the experiments increase with the number of discretization nodes in such a fashion that there are roughly 80 nodes for each wavelength. Tables 6–8 present the results for the kernels (6.1)–(6.3), respectively.

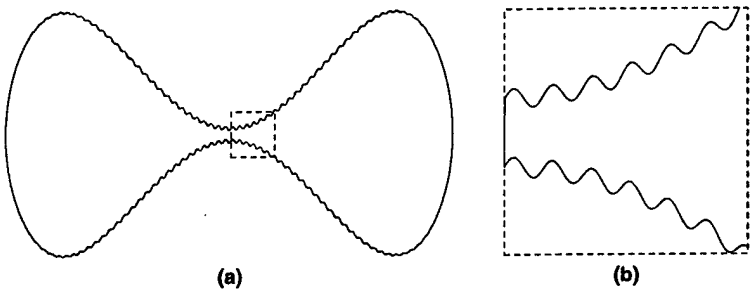


Fig. 6. (a) A rippled contour. (b) A close-up of the area marked by a dashed rectangle in (a). The horizontal axis of the contour has length 1 and the number of ripples change between the different experiments to keep a constant ratio of 80 discretization nodes per wavelength.

Table 6
Computational results for the double layer potential (6.1) associated with an exterior Laplace Dirichlet problem on the rippled contour shown in Fig. 6

N_{start}	N_{final}	t_{tot}	t_{solve}	E_{actual}	E_{res}	E_{pot}	c_{top}	σ_{min}	M
400	171	2.3e−01	1.0e−03	1.5e−10	1.1e−10	1.3e−07	7.4e+00	1.1e−01	1.5e+00
800	228	3.5e−01	1.0e−02	1.9e−10	1.3e−10	3.8e−08	9.7e+00	7.6e−02	3.0e+00
1600	306	7.3e−01	5.8e−03	1.3e−10	1.6e−10	5.5e−08	1.6e+01	5.2e−02	6.2e+00
3200	386	2.2e+00	8.5e−03	–	1.4e−10	7.5e−08	3.1e+01	3.9e−02	1.2e+01
6400	460	4.4e+00	1.7e−02	–	7.2e−11	8.2e−08	7.0e+01	3.3e−02	2.1e+01
12 800	536	9.6e+00	3.5e−02	–	5.9e−11	3.7e−08	1.4e+02	2.9e−02	4.0e+01
25 600	597	2.0e+01	7.6e−02	–	2.0e−11	1.4e−09	2.2e+02	–	7.6e+01
51 200	641	4.0e+01	1.5e−01	–	2.9e−11	–	6.2e+02	–	1.5e+02
102 400	688	(1.8e+01)	3.9e−01	–	1.2e−11	–	7.8e+02	–	2.9e+02

Table 7
Computational results for the single layer potential (6.2) associated with an exterior Laplace Dirichlet problem on the rippled contour shown in Fig. 6

N_{start}	N_{final}	t_{tot}	t_{solve}	E_{actual}	E_{res}	E_{pot}	c_{top}	σ_{min}	M
400	176	2.4e−01	9.2e−04	2.1e−09	1.7e−09	2.4e−05	1.6e+02	5.5e−04	1.6e+00
800	220	3.9e−01	3.8e−03	1.6e−08	3.0e−08	8.0e−06	1.1e+03	1.0e−05	3.1e+00
1600	256	6.9e−01	5.3e−03	5.2e−09	7.0e−09	9.8e−08	2.8e+02	1.6e−05	5.3e+00
3200	286	1.3e+00	7.6e−03	–	7.0e−09	1.6e−08	3.3e+02	1.2e−05	9.1e+00
6400	314	2.5e+00	1.4e−02	–	1.5e−07	2.3e−09	7.5e+02	2.1e−06	1.6e+01
12 800	342	4.6e+00	2.8e−02	–	2.4e−08	1.5e−09	4.7e+02	1.7e−07	2.9e+01
25 600	362	8.8e+00	6.2e−02	–	2.3e−08	2.2e−09	1.1e+03	9.7e−08	5.5e+01
51 200	374	1.7e+01	1.2e−01	–	2.1e−08	–	1.8e+03	3.1e−08	1.1e+02
102 400	386	(8.1e+0)	2.3e−01	–	1.5e−07	–	3.1e+03	–	2.1e+02

Table 8
Computational results for the kernel (6.3) associated with an exterior Helmholtz Dirichlet problem on the rippled contour shown in Fig. 6

k	N_{start}	N_{final}	t_{tot}	t_{solve}	E_{actual}	E_{res}	E_{pot}	c_{top}	σ_{min}	M
7	400	224	2.9e+00	9.0e−03	1.4e−07	6.9e−08	9.4e−07	1.2e+04	7.9e−01	3.2e+00
15	800	320	7.7e+00	1.9e−02	1.6e−07	7.4e−08	1.2e−07	3.9e+03	7.9e−01	8.2e+00
29	1600	470	2.1e+01	4.6e−02	–	6.7e−08	8.1e−08	7.4e+03	7.8e−01	2.0e+01
58	3200	704	6.1e+01	1.1e−01	–	5.2e−08	6.4e−08	1.2e+04	8.0e−01	5.0e+01
115	6400	1122	1.4e+02	2.9e−01	–	4.8e−08	7.5e−08	1.4e+04	8.0e−01	1.3e+02
230	12 800	1900	(4.7e+02)	(2.5e+01)	–	5.5e−08	7.5e−08	8.8e+04	8.0e−01	3.4e+02
461	25 600	3398	–	–	–	–	–	–	–	9.8e+02

The Helmholtz parameter k was chosen to keep the number of discretization points per wavelength constant at roughly 55 points per wavelength (resulting in a quadrature error about 10^{-12}). The times in parenthesis refer to experiments that did not fit in RAM.

We see that the asymptotic complexity of the algorithm remains essentially the same as for the smooth contour shown in Fig. 4. However, the constants involved are larger since more degrees of freedom are required to resolve the contour at the finest levels.

6.3. An interior problem close to a resonance

In this section we present results pertaining to *interior* Dirichlet problem on the contour shown in Fig. 7. While interior and exterior Laplace Dirichlet problems are quite similar in nature, the corresponding Helmholtz Dirichlet problems are fundamentally different in that the interior problem possesses resonances while the exterior does not. We will therefore focus exclusively on interior Helmholtz problems.

We present the results of two computational experiments, both relating to the Helmholtz kernel (6.3). In the first experiment, we scan a range of wave numbers k between 99.9 and 100.1. For each wave number, we computed the smallest singular value σ_{min} of the integral operator using the iteration technique described in Section 4.5. The resulting graph of σ_{min} versus k, shown in Fig. 8, clearly indicates the location of each resonance in this interval. The second experiment consists of factoring the inverse of the matrix corresponding to $k = 100.0110276\cdots$ for which $\sigma_{min} = 0.00001366\cdots$. The results, shown in Table 9, illustrate that the method does not experience any difficulty in factoring the inverse of a reasonably ill-conditioned matrix. In particular, the table shows that the factorization matrices $B^{(j)}$, $C^{(j)}$ and $D^{(j)}$, see (3.17), are well-conditioned.

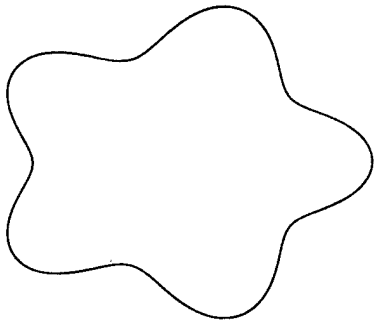


Fig. 7. A contour the shape of a smooth pentagram. Its diameter is 2.5 and its length is roughly 8.3.

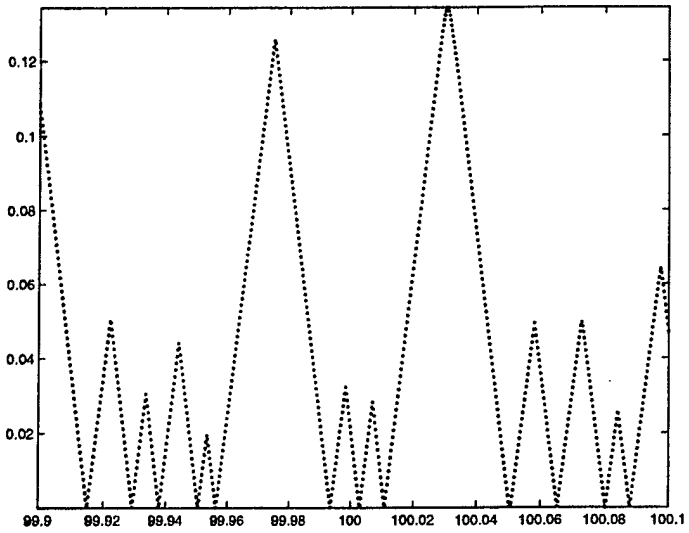


Fig. 8. Plot of $\sigma_{\min}$ versus k for an interior Helmholtz problem on the contour shown in Fig. 7. The values shown were computed using the iteration technique of Section 4.5 applied to a matrix of size $N = 6400$. Each point in the graph required about 60 s of CPU time.

Table 9
Details of the computation for the Helmholtz kernel (6.3) associated with an interior Dirichlet problem on the smooth pentagram shown in Fig. 7 for the case $N = 6400$ and $k = 100.011027569\cdots$

j	p_j	n_j	γ_j	t_j	$\|C^{(j)}\|_{\infty}$	$\|B^{(j)}\|_{\infty}$	$\|D^{(j)}\|_{\infty}$
1	128	50.00	0.76	15.50	1.12e+00	1.12e+00	4.20e−02
2	64	76.00	0.59	14.32	3.27e+01	3.27e+01	1.75e+00
3	32	89.72	0.60	8.94	1.63e+01	1.62e+01	9.28e−01
4	16	107.00	0.64	6.27	9.09e+00	9.17e+00	2.41e+00
5	8	138.00	0.72	5.97	7.32e+00	7.31e+00	3.64e+00
6	4	199.50	0.80	7.76	3.22e+00	3.23e+00	3.86e+00

For each level j, the table shows the number of clusters p_j on that level, the average size of a cluster n_j, the compression ratio γ_j, the time required for the factorization t_j and the size of the matrices $B^{(j)}$, $C^{(j)}$ and $D^{(j)}$ (see (3.17)) in the maximum norm. For this computation, $E_{\text{res}} = 2.8 \times 10^{-10}$, $E_{\text{pot}} = 3.3 \times 10^{-5}$ and $\sigma_{\min} = 1.4 \times 10^{-5}$.

6.4. A contour resembling an area integral

The final numerical experiment that we present is included to demonstrate that the efficiency of the factorization scheme deteriorates when it is applied to a curve for which the physical distance between two random points on the contour is not well predicted by their physical separation. One example of such a curve is the star-fish lattice illustrated in Fig. 9. Focusing on the double layer Laplace problem (6.1), we apply the factorization scheme to a matrix of size $N = 25\,600$ and compare the performance to that for the rippled dumb-bell shown in Fig. 6. Table 10 shows that the factorization of the matrix related to the starfish lattice took almost five times as long and resulted in a compressed matrix of over twice the size.

To understand the difference in performance between the different contours, we need to consider how the interaction rank of a cluster depends on its size. For the contours shown in Figs. 4, 6, and 7, we have seen that the rank of the interaction between a cluster of size m and the rest of the contour is effectively bounded by $\log m$. However, for the contour shown in Fig. 9 the corresponding bound is $\sqrt{m}$. Figs. 5 and 10 illus-

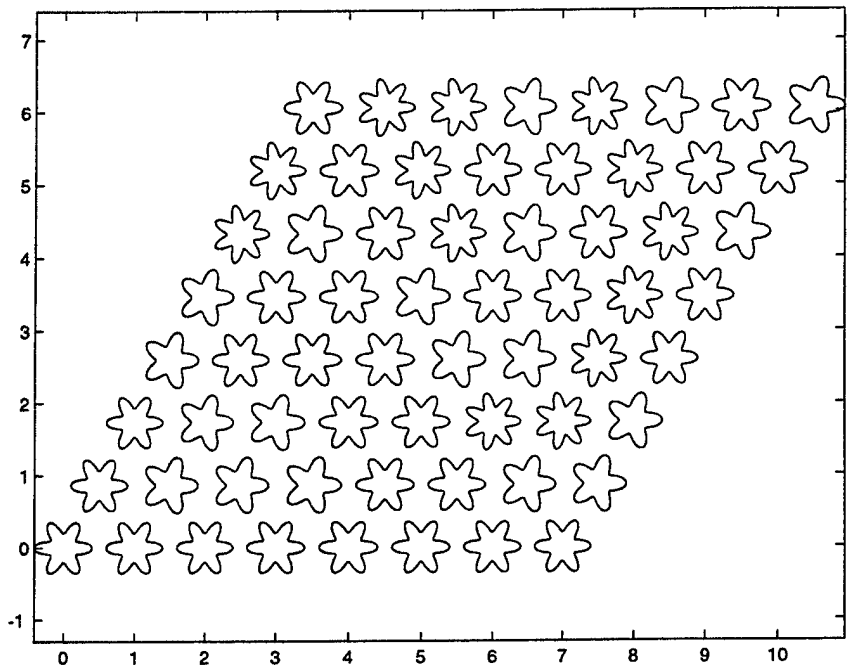


Fig. 9. The star-fish lattice contour; the physical distance between two random points on the contour is not well predicted by their distance along the contour.

Table 10
Test results for two experiments concerning the matrix obtained by discretizing the double layer Laplace problem (6.1)

Contour:	t_{tot}	N_{start}	N_{final}	M
Rippled dumb-bell	37s	25 600	559	86Mb
Star-fish lattice	172s	25 600	1202	210Mb

The table illustrates the difference in performance of the algorithm when applied to, on the one hand, the contour shown in Fig. 6 (top line), and on the other hand, the contour shown in Fig. 9 (lower line).

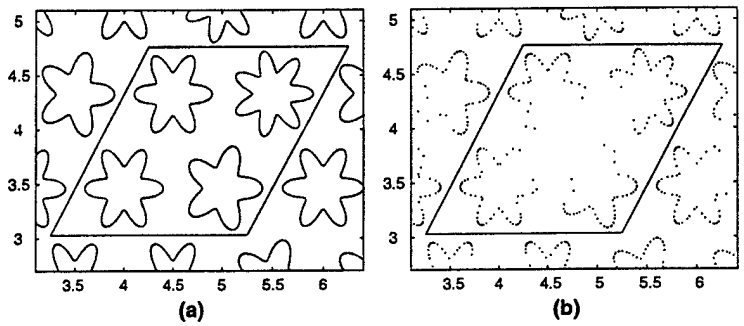


Fig. 10. Figure (a) shows a close-up of the star-fish lattice of Fig. 9. Figure (b) shows the nodes remaining after the interaction between the cluster formed by the points inside the parallelogram and the remainder of the contour has been compressed, cf. Fig. 5.

trate the difference. Thus, the asymptotic complexity of the scheme when applied to a contour similar to the star-fish lattice is $O(n^{3/2})$ rather than $O(n \log n)$.

7. Generalizations and conclusions

We have presented a numerical scheme that constructs a compressed factorization of the inverse of a matrix. The scheme is applicable to generic matrices whose off-diagonal blocks have rank-deficiencies but is most efficient when applied to matrices arising from the discretization of integral equations defined on one-dimensional contours. (Although such integral equations frequently arise in the analysis of boundary value problems in two dimensions, the dimension of the underlying space is of little relevance to the algorithm.) For equations with non-oscillatory kernels the computational complexity of the algorithm is $O(n \log^{\kappa} n)$ for most contours, where $\kappa = 1$ or 2, and n is the number of nodes in the discretization of the contour.

Comparing our implementations of the direct factorization scheme on the one hand and the FMM matrix-vector multiplication scheme on the other, we observed (i) that in a typical environment, the cost of constructing a factorization of the inverse is 15–20 times larger than the cost of a single FMM matrix-vector multiply, and (ii) that once the factorization of the inverse has been computed, the cost to apply it to a vector is 5–10 times smaller than the cost of a single FMM matrix-vector multiply. Thus, if an iterative solver requires less than 20 steps to converge, the iterative solver outperforms the direct solver for a single solve. However, if multiple right-hand sides are involved, the direct solver has a clear advantage. This observation is the foundation for [11].

Since the scheme is based on rank considerations only, it cannot work for boundary integral equations involving highly oscillatory kernels. However, since the interaction ranks are determined dynamically, the oscillation must be quite significant before the scheme becomes impracticable. Empirically, it was found that the scheme remains efficient for contours a couple of hundred wavelengths in size.

Another limitation of the scheme is that it does not achieve optimal efficiency when applied to a boundary integral equation set on either a contour similar to the one shown in Fig. 9, or on a two-dimensional surface. In either case, its computational complexity is $O(n^{3/2})$. Overcoming this limitation is a subject of on-going research.

Finally, we mention that the matrix factorization scheme presented in this paper can be modified to construct certain standard matrix factorizations (such as the singular value decomposition). This modification will be reported at a later date.

Acknowledgement

The first author was supported in part by the Office of Naval Research under contract #N00014-01-1-0364. The second author was supported in part by the Defense Advanced Research Projects Agency under contract # MDA972-00-1-0033.

References

[1] F.X. Canning, K. Rogivin, Fast direct solution of standard moment-method matrices, IEEE Antennas Propagation Mag. 40 (1998) 15–26.

[2] Yu. Chen, A fast direct algorithm for the Lippmann–Schwinger integral equation in two dimensions, Adv. Comput. Math. 16 (2-3) (2002) 175–190, Yu Chen, Modeling and computation in optics and electromagnetics. MR 2003b:65139.

[3] H. Cheng, Z. Gimbutas, P.G. Martinsson, and V. Rokhlin, On the compression of low rank matrices, Tech. report, Yale University, Department of Computer Science, 2003.

[4] W.C. Chew, An n^2 algorithm for the multiple scattering problem of n scatterers, Micro. Optical Tech. Lett. 2 (1989) 380–383.
[5] D. Gines, G. Beylkin, J. Dunn, LU factorization of non-standard forms and direct multiresolution solvers, Appl. Comput. Harmon. Anal. 5 (2) (1998) 156–201, MR99c:65087.
[6] Gene H. Golub, Charles F. Van Loan, Matrix computations, third ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996, MR 97g:65006.
[7] Ming Gu, Stanley C. Eisenstat, Efficient algorithms for computing a strong rank-revealing QR factorization, SIAM J. Sci. Comput. 17 (4) (1996) 848–869, MR 97h:65053.
[8] W. Hackbusch, A sparse matrix arithmetic based on H-matrices. I. Introduction to H-matrices, Computing 62 (2) (1999) 89–108, MR 2000c:65039.
[9] W. Hackbusch, S. Börm, Data-sparse approximation by adaptive H^2-matrices, Computing 69 (1) (2002) 1–35, MR 1 954 142.
[10] S. Kapur, V. Rokhlin, High-order corrected trapezoidal quadrature rules for singular functions, SIAM J. Numer. Anal. 34 (4) (1997) 1331–1356, MR 98k:65011.
[11] P.G. Martinsson, Fast evaluation of electro-static interactions in two phase dielectric media, Tech. report, Yale University, Dept. of Computer Science, 2004.
[12] E. Michielssen, A. Boag, A multilevel matrix decomposition algorithm for analysing scattering from large structures, IEEE Trans. Antennas and Propagation 44 (8) (1996) 1086–1093.
[13] E. Michielssen, A. Boag, W.C. Chew, Scattering from elongated objects: direct solution in $O(n \log^2 n)$ operations, IEEE Proc. H 143 (1996) 277–283.
[14] D. Scott, Analysis of the symmetric lanczos process, Tech. report, University of California at Berkeley, 1978.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 191 (2003) 40–74
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Second kind integral equations for the classical potential theory on open surfaces I: analytical apparatus

Shidong Jiang *,[1], Vladimir Rokhlin

Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA

Received 6 February 2003; accepted 21 May 2003

Abstract

A stable second kind integral equation formulation has been developed for the Dirichlet problem for the Laplace equation in two dimensions, with the boundary conditions specified on a collection of open curves. The performance of the obtained apparatus is illustrated with several numerical examples.

AMS: 65R10; 77C05

Keywords: Open surface problems; Laplace equation; Finite Hilbert transform; Second kind integral equation; Dirichlet problem

1. Introduction

Integral equations have been one of principal tools for the numerical solution of scattering problems for more than 30 years, both in the Helmholtz and Maxwell environments. Historically, most of the equations used have been of the first kind, since numerical instabilities associated with such equations have not been critically important for the relatively small-scale problems that could be handled at the time.

The combination of improved hardware with the recent progress in the design of "fast" algorithms has changed the situation dramatically. Condition numbers of systems of linear algebraic equations resulting from the discretization of integral equations of potential theory have become critical, and the simplest way to limit such condition numbers is by starting with second kind integral equations. Hence, interest has increased in reducing scattering problems to systems of *second kind* integral equations on the boundaries of the scatterers.

During the last several years, satisfactory integral equation formulations have been constructed in both acoustic (Helmholtz equation) and electromagnetic (Maxwell's equations) environments, whenever all of

* Corresponding author.
E-mail addresses: shidong@cs.yale.edu (S. Jiang), rokhlin@cs.yale.edu (V. Rokhlin).
[1] Supported in part by DARPA under Grant MDA972-00-1-0033, and by ONR under Grant N00014-01-1-0364.

0021-9991/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0021-9991(03)00304-8

the scattering surfaces are "closed" (i.e., scatterers have well-defined interiors, and have no infinitely thin parts). Boundary value problems for the biharmonic equation with boundary data specified on a collection of open curves have been investigated in some detail in [9–11]. However, a stable second kind integral equation formulation for scattering problems involving "open" surfaces does not appear to be present in the literature.

In this paper, we describe a stable second kind integral equation formulation for the Dirichlet problem for the Laplace equation in $\mathbb{R}^2$, with the boundary conditions specified on an "open" curve. We start with a detailed investigation of the case when the curve in question is the interval $[-1,1]$ on the real axis; then we generalize the obtained results for the case of (reasonably) general open curves.

The layout of the paper is as follows. In Section 2, the necessary mathematical and numerical preliminaries are introduced. Section 3 contains the exact statement of the problem. Section 4 contains an informal description of the procedure. In Sections 5 and 6, we investigate the cases of the straight line segment and of the general sufficiently smooth curve, respectively. In Section 7, we describe a simple numerical implementation of the scheme described in Section 6. The performance of the algorithm is illustrated in Section 8 with several numerical examples. Finally, in Section 9 we discuss several generalizations of the approach.

2. Analytical preliminaries

In this section, we summarize several results from classical and numerical analysis to be used in the remainder of the paper. Detailed references are given in the text.

2.1. Notation

Suppose that a,b are two real numbers with $a < b$, and $f,g : [a,b] \to \mathbb{C}$ is a pair of smooth functions, and that on the interval $[a,b]$, the function g has a single simple root s. Throughout this paper, we will be repeatedly encountering expressions of the form

$$\lim_{\epsilon \to 0} \left(\int_a^{s-\epsilon} \frac{f(t)}{g(t)}\,dt + \int_{s+\epsilon}^{b} \frac{f(t)}{g(t)}\,dt \right), \tag{1}$$

normally referred to as principal value integrals. In a mild abuse of notation, we will refer to expressions of the form (1) simply as integrals. We will also be fairly cavalier about the spaces on which operators of the type (1) operate; whenever the properties (smoothness, boundedness, etc.) required from a function are obvious from the context, their exact specifications are omitted.

2.2. Chebyshev polynomials and Chebyshev approximation

Chebyshev polynomials are frequently encountered in numerical analysis. As is well known, Chebyshev polynomials of the first kind $T_n : [-1,1] \to \mathbb{R} (n \geqslant 0)$ are defined by the formula

$$T_n(x) = \cos(n \arccos(x)) \tag{2}$$

and are orthogonal with respect to the inner product

$$(f,g) = \int_{-1}^{1} f(x) \cdot g(x) \cdot \frac{1}{\sqrt{1-x^2}}\,dx. \tag{3}$$

The Chebyshev nodes x_i of degree N are the zeros of T_N defined by the formula

$$x_i = \cos\frac{(2i+1)\pi}{2N}, \quad i = 0, 1, \ldots, N-1. \tag{4}$$

Chebyshev polynomials of the second kind $U_n : [-1,1] \to \mathbb{R} (n \geqslant 0)$ are defined by the formula

$$U_n(x) = \frac{\sin((n+1)\arccos(x))}{\sin(\arccos(x))} \tag{5}$$

and are orthogonal with respect to the inner product

$$(f,g) = \int_{-1}^{1} f(x) \cdot g(x) \cdot \sqrt{1-x^2}\,\mathrm{d}x. \tag{6}$$

The Chebyshev nodes of the second kind t_j of degree N are the zeros of U_N defined by the formula

$$t_j = \cos\frac{(N-j)\pi}{N+1}, \quad j = 0, 1, \ldots, N-1. \tag{7}$$

For a sufficiently smooth function $f : [-1,1] \to \mathbb{R}$, its Chebyshev expansion is defined by the formula

$$f(x) = \sum_{k=0}^{\infty} C_k \cdot T_k(x), \tag{8}$$

with the coefficients C_k given by the formulae

$$C_0 = \frac{1}{\pi}\int_{-1}^{1} f(x) \cdot T_0(x) \cdot (1-x^2)^{-1/2}\,\mathrm{d}x, \tag{9}$$

and

$$C_k = \frac{2}{\pi}\int_{-1}^{1} f(x) \cdot T_k(x) \cdot (1-x^2)^{-1/2}\,\mathrm{d}x, \tag{10}$$

for all $k \geqslant 1$. We will also denote by P_f^N the order $N-1$ Chebyshev approximation to the function f on the interval $[-1,1]$, i.e., the (unique) polynomial of order $N-1$ such that $P_f^N(x_i) = f(x_i)$ for all $i = 0, 1, \ldots, N-1$, with x_i the Chebyshev nodes defined by (4).

The following lemma provides an error estimate for the Chebyshev approximation (see, for example [5]).

Lemma 1. *If $f \in C^k[-1,1]$ (i.e., f has k continuous derivatives on the interval $[-1,1]$), then for any $x \in [-1,1]$,*

$$\left|P_f^N(x) - f(x)\right| = \mathrm{O}\left(\frac{1}{N^k}\right). \tag{11}$$

In particular, if f is infinitely differentiable, then the Chebyshev approximation converges superalgebraically (i.e., faster than any finite power of $1/N$ as $N \to \infty$).

2.3. The finite Hilbert transform

We will define the finite Hilbert transform $\widetilde{H}$ by the formula

$$\widetilde{H}(\varphi)(x) = \int_{-1}^{1} \frac{\varphi(t)}{t-x}\,\mathrm{d}t. \tag{12}$$

We then define the operator $\widetilde{K} : C^2[-1,1] \to L^2(-\infty,\infty)$ by the formula

$$\widetilde{K}(\varphi)(x) = \lim_{\epsilon \to 0} \left(\int_{-1}^{x-\epsilon} \frac{\varphi(t)}{(t-x)^2}\,dt + \int_{x+\epsilon}^{1} \frac{\varphi(t)}{(t-x)^2}\,dt - \frac{2\varphi(x)}{\epsilon} \right), \tag{13}$$

and observe that the limit (13) is often referred to as the finite part integral

$$\text{f.p.} \int_{-1}^{1} \frac{\varphi(t)}{(t-x)^2}\,dt \tag{14}$$

(see, for example Hadamard [8]).

The following theorem can be found in [13]; it provides sufficient conditions for the existence of the finite part integral (14), and establishes a connection between the finite Hilbert transform (12) and the finite part integral.

Theorem 2. *For any $\varphi \in C^2[-1,1]$, the limit* (13) *is a square-integrable function of x. Furthermore,*

$$\widetilde{K}(\varphi) = D \circ \widetilde{H}(\varphi), \tag{15}$$

with $D = \frac{d}{dx}$ the differentiation operator.

The following theorem (see, for example [21]) describes the inverse of the operator $\widetilde{H}$, to the extent that such an inverse exists

Theorem 3. *The null space of the operator $\widetilde{H}$ is spanned by the function $1/\sqrt{1-x^2}$. Furthermore, for any function $f \in L^p[-1,1]$ with $p > 1$, all solutions of the equation*

$$\widetilde{H}(\varphi) = f \tag{16}$$

are given by the formula

$$\varphi(x) = -\frac{1}{\pi^2} T^{-1} \circ \widetilde{H} \circ T(f)(x) + \frac{C}{\sqrt{1-x^2}}, \tag{17}$$

with C an arbitrary constant, and the operator $T : L^p[-1,1] \to L^p[-1,1]$ defined by the formula

$$T(f)(x) = \sqrt{1-x^2} \cdot f(x). \tag{18}$$

Applying Theorem 3 twice, we immediately obtain the following corollary:

Corollary 4. *For any $f \in C^1[-1,1]$, all solutions of the equation*

$$\widetilde{H} \circ \widetilde{H}(\varphi) = \widetilde{H}^2(\varphi) = f \tag{19}$$

are given by the formula

$$\varphi(x) = \frac{1}{\pi^4} T^{-1} \circ \widetilde{H}^2 \circ T(f)(x) + \frac{C_0}{\sqrt{1-x^2}} + \frac{C_1}{\sqrt{1-x^2}} \cdot \log\frac{1+x}{1-x}, \tag{20}$$

with C_0, C_1 two arbitrary constants.

2.4. Several elementary identities

In this section, we collect several identities from classical analysis to be used in the remainder of the paper. Lemma 5 states a well-known fact about the two-dimensional Poisson kernel $y/(x^2+y^2)$; it can be found in (for example) [19]. Lemma 6 provides explicit expressions for the finite Hilbert transform operating on Chebyshev polynomials, where (22) is a direct consequence of Lemma 3, and (23), (24) can be found in [2]. Lemma 7 lists several standard definite integrals; all can be found (in a somewhat different form) in [6]. Finally, Lemma 8 follows from the definition of curvature $c(t)$ found in elementary differential geometry (cf. [3]).

Lemma 5. *Suppose that* $\sigma \in L^p[-1,1]$ $(p \geqslant 1)$. *Then*

$$\lim_{y\to 0} \int_{-1}^{1} \frac{|y|}{\pi((x-t)^2+y^2)} \cdot \sigma(t)\,dt = \sigma(x), \tag{21}$$

for almost all $x \in [-1,1]$.

Lemma 6. *For any* $x \in (-1,1)$,

$$\int_{-1}^{1} \frac{1}{t-x} \cdot \frac{1}{\sqrt{1-t^2}}\,dt = 0, \tag{22}$$

and

$$\int_{-1}^{1} \frac{\sqrt{1-t^2}}{t-x} \cdot U_{n-1}(t)\,dt = -\pi \cdot T_n(x), \tag{23}$$

$$\int_{-1}^{1} \frac{1}{t-x} \cdot \frac{1}{\sqrt{1-t^2}} \cdot T_n(t)\,dt = \pi \cdot U_{n-1}(x), \tag{24}$$

for any $n \geqslant 1$.

Lemma 7.

1. *For any* $x,t \in (-1,1)$ *and* $x \neq t$,

$$\int_{-1}^{1} \frac{1}{(s-x)(s-t)}\,ds = \frac{\log\frac{1-x}{1-t} - \log\frac{1+x}{1+t}}{x-t}. \tag{25}$$

2. *For any* $(x,y) \in \mathbb{R}^2 \setminus [-1,1]$ *and* $t \in (-1,1)$,

$$\int_{-1}^{1} \frac{(s-x)}{((s-x)^2+y^2)(s-t)}\,ds = \frac{|y| \cdot \left(\arctan\left(\frac{1-x}{|y|}\right) + \arctan\left(\frac{1+x}{|y|}\right)\right)}{((x-t)^2+y^2)} + \frac{(x-t) \cdot \left(\log\frac{(1-x)^2+y^2}{(1-t)^2} - \log\frac{(1+x)^2+y^2}{(1+t)^2}\right)}{2((x-t)^2+y^2)}. \tag{26}$$

3. *For any* $x \in (-1,1)$,

$$\int_{-1}^{1} \log|x-t| \cdot \frac{1}{\sqrt{1-t^2}}\,dt = -\pi \cdot \log 2, \tag{27}$$

$$\int_{-1}^{1} \log|x-t| \cdot \frac{t}{\sqrt{1-t^2}}\,\mathrm{d}t = -\pi \cdot x, \tag{28}$$

$$\int_{-1}^{1} \frac{1}{t-x} \cdot \frac{1}{\sqrt{1-t^2}} \cdot \log(1+t)\,\mathrm{d}t = \frac{\pi \cdot \arccos x}{\sqrt{1-x^2}}, \tag{29}$$

$$\int_{-1}^{1} \frac{1}{t-x} \cdot \frac{1}{\sqrt{1-t^2}} \cdot \log(1-t)\,\mathrm{d}t = \frac{\pi \cdot (\arccos(x)-\pi)}{\sqrt{1-x^2}}, \tag{30}$$

$$\int_{-1}^{1} \frac{\sqrt{1-t^2}}{t-x} \cdot \log(1+t)\,\mathrm{d}t = \pi \cdot (\arccos(x) \cdot \sqrt{1-x^2} + \log(2) \cdot x - 1), \tag{31}$$

$$\int_{-1}^{1} \frac{\sqrt{1-t^2}}{t-x} \cdot \log(1-t)\,\mathrm{d}t = \pi \cdot ((\arccos(x)-\pi) \cdot \sqrt{1-x^2} + \log(2) \cdot x + 1). \tag{32}$$

Lemma 8. *Suppose that $\gamma : [0, L] \to \mathbb{R}^2$ is a sufficiently smooth curve parametrized by its arc length with the unit normal and the unit tangent vectors at $\gamma(t)$ denoted by $N(t)$ and $T(t)$, respectively. Suppose further that the function $u : \mathbb{R}^2 \to \mathbb{R}$ is twice continuously differentiable. Then at the point $\gamma(t)$, the Laplacian of u is given by the formula*

$$\Delta u = N \cdot \nabla\nabla u \cdot N - c(t) N \cdot \nabla u + \frac{\mathrm{d}^2}{\mathrm{d}t^2} u(\gamma(t)) = \frac{\partial^2 u}{\partial N(t)^2} - c(t) \cdot \frac{\partial u}{\partial N(t)} + \frac{\partial^2 u}{\partial T(t)^2}, \tag{33}$$

where the curvature $c(t)$ at $\gamma(t)$ is defined by $\mathrm{d}^2\gamma/\mathrm{d}t^2 = c(t)N(t)$.

2.5. The Poincaré–Bertrand formula

For a fixed point $x \in (-1, 1)$, we will consider two repeated integrals

$$A = \int_{-1}^{1} \frac{\varphi_1(t)}{t-x} \cdot \left(\int_{-1}^{1} \frac{\varphi_2(s)}{s-t}\,\mathrm{d}s \right) \mathrm{d}t, \tag{34}$$

$$B = \int_{-1}^{1} \varphi_2(s) \cdot \left(\int_{-1}^{1} \frac{\varphi_1(t)}{(t-x)(s-t)}\,\mathrm{d}t \right) \mathrm{d}s, \tag{35}$$

differing from each other only in the order of integration. Both integrals exist almost everywhere for a fairly broad class of functions. However, they are not, in general, equal to one another. The following lemma establishes the connection between them (see, for example [17,21]; the result is usually referred to as the Poincaré–Bertrand formula.

Lemma 9. *Suppose that $\varphi_1 \in L^p[-1, 1]$, $\varphi_2 \in L^q[-1, 1]$. Then if*

$$\frac{1}{p} + \frac{1}{q} < 1, \tag{36}$$

then

$$\int_{-1}^{1} \frac{\varphi_1(t)}{t-x} \cdot \left(\int_{-1}^{1} \frac{\varphi_2(s)}{s-t} \, \mathrm{d}s \right) \mathrm{d}t = -\pi^2 \cdot \varphi_1(x) \cdot \varphi_2(x) + \int_{-1}^{1} \varphi_2(s) \cdot \left(\int_{-1}^{1} \frac{\varphi_1(t)}{(t-x)(s-t)} \, \mathrm{d}t \right) \mathrm{d}s \tag{37}$$

for almost all $x \in (-1,1)$.

2.6. Potential theory

In this section, we introduce some terminology standard in potential theory and state several technical lemmas to be used subsequently. We will define the potential $G_{x_0} : \mathbb{R}^2 \setminus \{x_0\} \to \mathbb{R}$ of a unit charge located at the point $x_0 \in \mathbb{R}^2$ by the formula

$$G_{x_0}(x) = \log(\|x - x_0\|). \tag{38}$$

Suppose that $\gamma : [0,L] \to \mathbb{R}^2$ is a sufficiently smooth curve parametrized by its arc length, and that γ is an open curve (i.e., $\gamma(0) \neq \gamma(L)$). The image of γ will be denoted by Γ, and the unit normal and the unit tangent vectors to γ at the point $\gamma(t)$ will be denoted by $N(t)$ and $T(t)$, respectively. Given an integrable function $\sigma : [0,L] \to \mathbb{R}$, we will refer to the functions $S_{\gamma,\sigma} : \mathbb{R}^2 \to \mathbb{R}$ and $D_{\gamma,\sigma}, Q_{\gamma,\sigma} : \mathbb{R}^2 \setminus \Gamma \to \mathbb{R}$, defined by the formulae

$$S_{\gamma,\sigma}(x) = \int_0^L G_{\gamma(t)}(x) \cdot \sigma(t) \, \mathrm{d}t, \tag{39}$$

$$D_{\gamma,\sigma}(x) = \int_0^L \frac{\partial G_{\gamma(t)}(x)}{\partial N(t)} \cdot \sigma(t) \, \mathrm{d}t, \tag{40}$$

$$Q_{\gamma,\sigma}(x) = \int_0^L \frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2} \cdot \sigma(t) \, \mathrm{d}t, \tag{41}$$

as the single, double, and quadruple layer potentials, respectively.

The functions $(\partial G_{\gamma(t)}(x))/(\partial N(t)), (\partial^2 G_{\gamma(t)}(x))/(\partial N(t)^2) : \mathbb{R}^2 \setminus \gamma(t) \to \mathbb{R}$ are often referred to as the dipole and quadrupole potentials, respectively. Obviously,

$$\frac{\partial G_{\gamma(t)}(x)}{\partial N(t)} = -\frac{\langle N(t), x - \gamma(t) \rangle}{\|x - \gamma(t)\|^2}, \tag{42}$$

$$\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2} = -\frac{2\langle N(t), x - \gamma(t) \rangle^2}{\|x - \gamma(t)\|^4} + \frac{1}{\|x - \gamma(t)\|^2}. \tag{43}$$

In particular, if γ is a straight line segment $I_L = [0,L]$ on the real axis, then

$$\frac{\partial G_{I(s+t)}(I(s) - h \cdot N(s))}{\partial N(s+t)} = \frac{h}{h^2 + t^2}, \tag{44}$$

$$\frac{\partial^2 G_{I(s+t)}(I(s) - h \cdot N(s))}{\partial N(s+t)^2} = \frac{t^2 - h^2}{(h^2 + t^2)^2}. \tag{45}$$

The following two lemmas can be found in [14]. Lemma 10 states a standard fact from elementary differential geometry of curves; Lemma 11 describes the local behavior on a curve of the potential of a quadrupole located on that curve and oriented normally to it.

Lemma 10. *Suppose that $\gamma : [0,L] \to \mathbb{R}^2$ is a sufficiently smooth curve parametrized by its arc length with the unit normal and the unit tangent vectors at $\gamma(t)$ denoted by $N(t)$ and $T(t)$, respectively. Then, there exist a positive real number β (dependent on γ), and two continuously differentiable functions $f, g : (-\beta,\beta) \to \mathbb{R}$ (dependent on γ), such that for any $t \in [0,L]$,*

$$\gamma(t+s) - \gamma(t) = \left(s - \frac{c(t)^2 \cdot s^3}{6} + s^4 \cdot f(s)\right) \cdot T(t) + \left(\frac{c(t) \cdot s^2}{2} + s^3 \cdot g(s)\right) \cdot N(t), \tag{46}$$

for all $s \in (-\beta,\beta)$, where $c(t)$ in (46) is the curvature of γ at the point $\gamma(t)$.

Lemma 11. *Suppose that $\gamma : [0,L] \to \mathbb{R}^2$ is a sufficiently smooth curve parametrized by its arc length. Then, there exist real positive numbers A, β, h_0 such that for any $s \in [0,L]$,*

$$\left| \frac{\partial^2 G_{\gamma(s+t)}(\gamma(s) - h \cdot N(s))}{\partial N(s+t)^2} - \frac{t^2 - h^2}{(h^2+t^2)^2} - \frac{c \cdot h \cdot t^2 \cdot (5h^2 + t^2)}{(h^2+t^2)^3} \right| \leqslant A, \tag{47}$$

for all $t \in (-\beta,\beta)$, $0 \leqslant h < h_0$, where the coefficient c in (47) is the positive curvature of γ at the point $\gamma(s)$.

Similarly, the following lemma describes the local behavior on a curve of the potential of a dipole located on that curve and oriented normally to it; it also describes the local behavior on a curve of the tangential derivative of the potential of a charge located on that curve. Its proof is virtually identical to that of Lemma 11.

Lemma 12. *Under the conditions of Lemma 11, there exist real positive numbers A, β, h_0 such that for any $s \in [0,L]$,*

$$\left| \frac{\partial G_{\gamma(s+t)}(\gamma(s) - h \cdot N(s))}{\partial N(s+t)} - \frac{h}{h^2+t^2} \right| \leqslant A, \tag{48}$$

$$\left| \frac{\partial G_{\gamma(s+t)}(\gamma(s) - h \cdot N(s))}{\partial T(s+t)} - \frac{t}{h^2+t^2} \right| \leqslant A, \tag{49}$$

for all $t \in (-\beta,\beta)$, $0 \leqslant h < h_0$.

We will define the function $M_{\gamma,\sigma} : \mathbb{R}^2 \setminus \Gamma \to \mathbb{R}$ by the formula

$$M_{\gamma,\sigma}(x) = Q_{\gamma,\sigma}(x) - D_{\gamma,c\sigma}(x) = \int_0^L \left(\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2} - c(t) \cdot \frac{\partial G_{\gamma(t)}(x)}{\partial N(t)} \right) \cdot \sigma(t)\,\mathrm{d}t, \tag{50}$$

for all $x \in \mathbb{R}^2 \setminus \Gamma$ and observe that $M_{\gamma,\sigma}$ is the difference of a quadruple layer potential and a weighted double layer potential with the weight equal to the curvature $c(t)$. The following theorem is a direct consequence of Lemmas 11 and 12; it states that under certain conditions the function $M_{\gamma,\sigma}$ defined by (50) can be continuously extended to the whole plane $\mathbb{R}^2$.

Theorem 13. *Suppose that $\gamma : [0,L] \to \mathbb{R}^2$ is a sufficiently smooth open curve parametrized by its arc length, and that $\sigma : [0,L] \to \mathbb{R}$ is a function continuous on $[0,L]$, whose second derivative is continuous on $(0,L)$. Then the function $M_{\gamma,\sigma}$ can be continuously extended to $\mathbb{R}^2 \setminus \{\gamma(0), \gamma(L)\}$ with the limiting value on $\gamma(0,L)$ defined by the formula*

$$M_{\gamma,\sigma}(\gamma(x)) = \text{f.p.} \int_0^L \frac{\partial^2 G_{\gamma(t)}(\gamma(x))}{\partial N(t)^2} \cdot \sigma(t)\,dt - \int_0^L c(t) \cdot \frac{\partial G_{\gamma(t)}(\gamma(x))}{\partial N(t)} \cdot \sigma(t)\,dt \tag{51}$$

for all $x \in (0,L)$. Furthermore, if σ satisfies the additional condition that

$$|\sigma(x)| \leqslant C \cdot (x \cdot (L-x))^{\alpha}, \tag{52}$$

with some $C > 0$, $\alpha > 1$ for all $x \in [0,L]$, then $M_{\gamma,\sigma}$ can be further continuously extended to $\mathbb{R}^2$ with the limiting values on $\gamma(0)$, $\gamma(L)$ given by the improper integrals

$$M_{\gamma,\sigma}(\gamma(0)) = \int_0^L \left(\frac{\partial^2 G_{\gamma(t)}(\gamma(0))}{\partial N(t)^2} - c(t) \cdot \frac{\partial G_{\gamma(t)}(\gamma(0))}{\partial N(t)} \right) \cdot \sigma(t)\,dt, \tag{53}$$

$$M_{\gamma,\sigma}(\gamma(L)) = \int_0^L \left(\frac{\partial^2 G_{\gamma(t)}(\gamma(L))}{\partial N(t)^2} - c(t) \cdot \frac{\partial G_{\gamma(t)}(\gamma(L))}{\partial N(t)} \right) \cdot \sigma(t)\,dt, \tag{54}$$

respectively.

Definition 14. We will denote by E the linear subspace of $C[0,L]$, consisting of functions σ satisfying the following two conditions:

1. σ is twice continuously differentiable on $(0,L)$;
2. σ satisfies the condition (52).

We then define the integral operator $M_\gamma : E \to C[0,L]$ via the formula

$$M_\gamma(\sigma)(x) = M_{\gamma,\sigma}(\gamma(x)). \tag{55}$$

The following lemma states that the operator M_γ on a sufficiently smooth open curve γ is a compact perturbation of the same operator M_{I_L} on the line segment $I_L = [0,L]$.

Lemma 15. *Suppose that $\gamma : [0,L] \to \mathbb{R}^2$ is a sufficiently smooth open curve parametrized by its arc length. Suppose further that the operator $R_\gamma : C[0,L] \to C[0,L]$ is defined by the formula*

$$R_\gamma(\sigma)(x) = \int_0^L r(x,t) \cdot \sigma(t)\,dt \tag{56}$$

with the function $r : [0,L] \times [0,L] \to \mathbb{R}$ defined by the formula

$$r(x,t) = \frac{\partial^2 G_{\gamma(t)}(\gamma(x))}{\partial N(t)^2} - c(t) \cdot \frac{\partial G_{\gamma(t)}(\gamma(x))}{\partial N(t)} - \frac{\partial^2 G_{I_L(t)}(x)}{\partial N(t)^2} \tag{57}$$

for all $x \neq t$, and by the formula

$$r(t,t) = \frac{c(t)^2}{12} \tag{58}$$

for all $x = t$, with $c(t)$ denoting the curvature of γ at the point $\gamma(t)$. Then

$$r(x,t) = -\frac{2\langle N(t), \gamma(x) - \gamma(t)\rangle^2}{\|\gamma(x) - \gamma(t)\|^4} + \frac{1}{\|\gamma(x) - \gamma(t)\|^2} + c(t) \cdot \frac{\langle N(t), \gamma(x) - \gamma(t)\rangle}{\|\gamma(x) - \gamma(t)\|^2} - \frac{1}{(x-t)^2} \tag{59}$$

for all $x \neq t$. Furthermore, r is continuous on $[0,L] \times [0,L]$, so that the operator R_γ is compact. Finally, if $\sigma \in E$ (see Definition 14 above), then

$$M_\gamma(\sigma)(x) = M_{I_L}(\sigma)(x) + R_\gamma(\sigma)(x). \tag{60}$$

Proof. Eq. (60) follows directly from the combination of (51), (56), (57) and the fact that the curvature is zero everywhere on the line segment I_L. (59) is a direct consequence of (42), (43), (45), (57). In order to prove that r is continuous on $[0,L] \times [0,L]$, we start with observing that since $\gamma \in C^2[0,L]$, it is sufficient to demonstrate that

$$\lim_{s\to 0} r(t+s,t) = \frac{c(t)^2}{12}. \tag{61}$$

Replacing x in (59) with $t+s$, we obtain

$$r(t+s,t) = -\frac{2\langle N(t), \gamma(t+s)-\gamma(t)\rangle^2}{\|\gamma(t+s)-\gamma(t)\|^4} + \frac{1}{\|\gamma(t+s)-\gamma(t)\|^2} + c(t)\cdot\frac{\langle N(t), \gamma(t+s)-\gamma(t)\rangle}{\|\gamma(t+s)-\gamma(t)\|^2} - \frac{1}{s^2}. \tag{62}$$

Substituting (46) into (62), we have

$$r(t+s,t) = -\frac{2p(s)^2}{d(s)^2} + c(t)\cdot\frac{p(s)}{d(s)} + \frac{1-d(s)}{s^2\cdot d(s)}, \tag{63}$$

where the functions $p, d : (-\beta,\beta) \to \mathbb{R}$ are given be the formulae

$$p(s) = \frac{c(t)}{2} + s\cdot g(s), \tag{64}$$

$$d(s) = \left(1 - \frac{c(t)^2\cdot s^2}{6} + s^3\cdot f(s)\right)^2 + \left(\frac{c(t)\cdot s}{2} + s^2\cdot g(s)\right)^2, \tag{65}$$

with β a positive real number, and the functions f, g provided by Lemma 10. Since f, g are continuously differentiable on $(-\beta,\beta)$ (see Lemma 10), we have

$$\lim_{s\to 0}\frac{p(s)}{d(s)} = \frac{c(t)}{2}, \tag{66}$$

$$\lim_{s\to 0}\frac{1-d(s)}{s^2\cdot d(s)} = \frac{c(t)^2}{12}. \tag{67}$$

Now, we obtain (61) by substituting (66), (67) into (63). □

Remark 16. A somewhat involved analysis shows that for any $k \geqslant 1$ and $\gamma \in C^{k+2}[0,L]$, the function r (see (57) above) is k times continuously differentiable. The proof of this fact is technical, and the fact itself is peripheral to the purpose of this paper; thus, the proof is omitted.

3. The exact statement of the problem

Suppose that γ is a sufficiently smooth open curve, and that the image of γ is denoted by Γ. We will denote by S_γ the set of continuous functions on $\mathbb{R}^2$ with continuous second derivatives in the complement of Γ, i.e.,

$$S_\gamma = C^2(\mathbb{R}^2 \setminus \Gamma) \cap C(\mathbb{R}^2). \tag{68}$$

We will consider the Dirichlet problem for the Laplace equation in $\mathbb{R}^2$, with the boundary conditions specified on γ:

Given a function $f : \Gamma \to \mathbb{R}$, find a bounded solution $u \in S_\gamma$ to the Laplace equation

$$\Delta u = 0 \quad \text{in } \mathbb{R}^2 \setminus \Gamma \tag{69}$$

satisfying the Dirichlet boundary condition

$$u = f \quad \text{on } \Gamma. \tag{70}$$

The following theorem can be found in [15].

Theorem 17. *If $f \in C^2(\Gamma)$, then there exists a unique bounded solution in S_γ to the problem* (69) *and* (70).

Remark 18. Certain physical problems lead to modifications of the problem (69) and (70). For example, the boundedness of the solution at infinity might be replaced with logarithmic growth, the boundary might consist of several disjoint components, etc. Extensions of Theorem 17 to these cases are straightforward, and can be found, for example, in [17].

4. Analytical apparatus I: informal description

In this section, we will present an informal description of the procedure. We assume that $\gamma : [-1, 1] \to \mathbb{R}^2$ is a sufficiently smooth "open" (i.e., $\gamma(-1) \neq \gamma(1)$) curve with the parametrization

$$\gamma(t) = \tilde{\gamma}\left(\frac{L}{2} \cdot (t+1)\right), \tag{71}$$

where L is the total arc length of the curve, and $\tilde{\gamma} : [0, L] \to \mathbb{R}^2$ is the same curve parametrized by its arc length. The image of γ will be denoted by Γ. We start with observing that the solution u of the Dirichlet problem (69) and (70) must satisfy the following four conditions:

(a) u is harmonic in $\mathbb{R}^2 \setminus \Gamma$;
(b) u is bounded at infinity;
(c) u is continuous across Γ;
(d) u is equal to the prescribed data f on Γ.

Our goal is to construct a second kind integral formulation for the Dirichlet problem (69) and (70). Standard approaches in classical potential theory call for representing u in $\mathbb{R}^2 \setminus \Gamma$ via single or double layer potentials so that conditions (a), (b) are automatically satisfied, and conditions (c), (d) lead to a boundary integral equation via the so-called jump relations of single and double layer potentials (see, for example [16]). However, in the case of an open curve, if u is represented via a double layer potential, the condition (c) cannot be satisfied since any non-zero double layer potential has a jump across the boundary; and if u is represented via a single layer potential, while the single layer potential can be continuously extended across the boundary, the condition (d) will lead to an integral equation of the *first* kind. Hence, classical tools of potential theory turn out to be insufficient for dealing with open surface problems.

It is shown in [14] that the quadruple layer potential has a jump across the boundary which is proportional to the curvature of the curve. Combining this observation with the well-known fact that the double layer potential has a jump across the boundary which is independent of the curvature, we observe that the sum of a quadruple layer potential and a weighted double layer potential with the weight equal to the curvature given by the formula

$$\int_{-1}^{1}\left(\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2}-c(t)\cdot\frac{\partial G_{\gamma(t)}(x)}{\partial N(t)}\right)\cdot\sigma(t)\,\mathrm{d}t \tag{72}$$

can be continuously extended across the boundary. However, if u is represented via (72), then the condition (d) will lead to a *hypersingular* integral equation. It is also shown in [14] that the product of the hypersingular integral operator with the single layer potential operator is a second kind integral operator in the case of a closed boundary. Thus, one is naturally lead to consider the operator P_γ defined by the formula

$$P_\gamma(\sigma)(x)=\int_{-1}^{1}\left(\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2}-c(t)\cdot\frac{\partial G_{\gamma(t)}(x)}{\partial N(t)}\right)\cdot\left(\int_{-1}^{1}\log|t-s|\cdot\sigma(s)\,\mathrm{d}s\right)\mathrm{d}t. \tag{73}$$

Obviously, $P_\gamma(\sigma)$ is not defined when $x\in\Gamma$, and we will define the operator B_γ by the formula

$$B_\gamma(\sigma)(t)=\lim_{x\to\gamma(t)}P_\gamma(\sigma)(x). \tag{74}$$

In the special case when γ is the interval $I=[-1,1]$ on the real axis, (73) assumes the form

$$P_I(\sigma)(x,y)=\frac{1}{2}\int_{-1}^{1}\frac{\partial^2}{\partial y^2}\log((x-s)^2+y^2)\cdot\left(\int_{-1}^{1}\log|s-t|\cdot\sigma(t)\,\mathrm{d}t\right)\mathrm{d}s, \tag{75}$$

and the operator B_I is defined by the formula

$$B_I(\sigma)(x)=\lim_{y\to 0}P_I(\sigma)(x,y). \tag{76}$$

The operator B_I turns out to have a remarkably simple analytical structure (see Section 5.4 below); its natural domain consists of functions of the form

$$\frac{1}{\sqrt{1-x^2}}\cdot\varphi(x)+\frac{1}{\sqrt{1-x^2}}\cdot\log\frac{1+x}{1-x}\cdot\psi(x), \tag{77}$$

with φ,ψ smooth functions, and when restricted to functions of the form (77), it has a null-space of dimension 2, spanned by the functions

$$\frac{1}{\sqrt{1-x^2}}, \tag{78}$$

$$\frac{1}{\sqrt{1-x^2}}\cdot\log\frac{1+x}{1-x}. \tag{79}$$

In Section 5.4, we construct a generalized (in the appropriate sense) inverse of B_I; in a mild abuse of notation, we will refer to it as B_I^{-1}.

Now, if we represent the solution of the Problem (69) and (70) in the form

$$u(x)=P_\gamma(\sigma)(x), \tag{80}$$

then the conditions (c) and (d) will lead to the equation

$$B_\gamma(\sigma)(t)=f(t), \tag{81}$$

with σ the unknown density. It turns out that (81) behaves *almost* like an integral equation of the second kind; the only problem is that the kernel of B_γ is strongly singular at the ends. Fortunately, the operator

$$\widetilde{B}_\gamma = B_\gamma \circ B_I^{-1}, \tag{82}$$

restricted to smooth functions, *is* a sum of the identity and a compact operator. In other words, $\widetilde{B}_\gamma$ is a *second kind* integral operator. Therefore, our representation for the solution of the Problem (69) and (70) takes the form

$$u(x) = \widetilde{P}_\gamma(\eta)(x) = P_\gamma \circ B_I^{-1}(\eta)(x), \tag{83}$$

with η the solution of the integral equation

$$\widetilde{B}_\gamma(\eta)(t) = f(t). \tag{84}$$

Finally, we remark that minor complications arise from the non-uniqueness of B_I^{-1} (see (78) and (79) above); they are resolved in Section 6.3.

5. Analytical apparatus II: open surface problem for the line segment $I=[-1,1]$

5.1. The integral operator P_I

Definition 19. We will denote by F_I the set of functions $\sigma : (-1,1) \to \mathbb{R}$ of the form

$$\sigma(x) = \frac{1}{\sqrt{1-x^2}} \cdot \varphi(x) + \frac{1}{\sqrt{1-x^2}} \cdot \log\frac{1+x}{1-x} \cdot \psi(x), \tag{85}$$

with $\varphi, \psi : [-1,1] \to \mathbb{R}$ twice continuously differentiable, and satisfying the conditions

$$\int_{-1}^{1} \log|1+t| \cdot \sigma(t)\,dt = 0, \tag{86}$$

$$\int_{-1}^{1} \log|1-t| \cdot \sigma(t)\,dt = 0. \tag{87}$$

We will consider the integral operator $P_I : F_I \to C^2(\mathbb{R}^2 \setminus I)$ defined by the formula

$$P_I(\sigma)(x,y) = \int_{-1}^{1} K_I(x,y,t) \cdot \sigma(t)\,dt = \frac{1}{2}\int_{-1}^{1} \frac{\partial^2}{\partial y^2} \log((x-s)^2+y^2) \cdot \left(\int_{-1}^{1} \log|s-t| \cdot \sigma(t)\,dt\right) ds. \tag{88}$$

Obviously, P_I converts a function $\sigma \in F_I$ into a quadruple layer potential whose density $D(\sigma)$ is in turn represented by a single layer potential

$$D(\sigma)(x) = \int_{-1}^{1} \log|x-t| \cdot \sigma(t)\,dt. \tag{89}$$

The following lemma provides an explicit expression for the kernel K_I of P_I.

Lemma 20. *For any* $\sigma \in F_I$,

$$K_I(x,y,t) = \frac{|y| \cdot \left(\arctan\left(\frac{1-x}{|y|}\right) + \arctan\left(\frac{1+x}{|y|}\right)\right)}{((x-t)^2+y^2)} + \frac{(x-t) \cdot \left(\log\frac{(1-x)^2+y^2}{(1-t)^2} - \log\frac{(1+x)^2+y^2}{(1+t)^2}\right)}{2((x-t)^2+y^2)}, \tag{90}$$

for any $(x,y) \in \mathbb{R}^2 \setminus I$ *and any* $t \in (-1,1)$.

Proof. Since $\log((x-s)^2+y^2)$ satisfies the Laplace equation for any $(x,y) \neq (s,0)$, we have

$$\frac{\partial^2}{\partial y^2}\log((x-s)^2+y^2) = -\frac{\partial^2}{\partial s^2}\log((x-s)^2+y^2); \tag{91}$$

substituting (91) into (88) and integrating by parts once, we obtain

$$P_I(\sigma)(x,y) = \frac{1}{2}\int_{-1}^{1}\frac{\partial}{\partial s}\log((x-s)^2+y^2)\cdot\left(\int_{-1}^{1}\frac{\partial}{\partial s}\log|s-t|\cdot\sigma(t)\,dt\right)ds$$
$$-\frac{(1-x)}{(x-1)^2+y^2}\cdot\int_{-1}^{1}\log|1-t|\cdot\sigma(t)\,dt - \frac{(1+x)}{(x+1)^2+y^2}\cdot\int_{-1}^{1}\log|1+t|\cdot\sigma(t)\,dt. \tag{92}$$

Combining (92) with (86), (87) and changing the order of integration, we have

$$P_I(\sigma)(x,y) = \int_{-1}^{1}\left(\frac{1}{2}\int_{-1}^{1}\frac{\partial}{\partial s}\log((s-x)^2+y^2)\cdot\frac{\partial}{\partial s}\log|s-t|\,ds\right)\cdot\sigma(t)\,dt. \tag{93}$$

Hence,

$$K_I(x,y,t) = \frac{1}{2}\int_{-1}^{1}\frac{\partial}{\partial s}\log((s-x)^2+y^2)\cdot\frac{\partial}{\partial s}\log|s-t|\,ds = \int_{-1}^{1}\frac{(s-x)}{((s-x)^2+y^2)(s-t)}\,ds. \tag{94}$$

Now, (90) follows immediately from the combination of (26) and (94). □

5.2. The boundary integral operator B_I

We will define the integral operator $B_I : F_I \to L^1[-1,1]$ (see (85)) by the formula

$$B_I(\sigma)(x) = \lim_{y\to 0} P_I(\sigma)(x,y) = \lim_{y\to 0}\int_{-1}^{1} K_I(x,y,t)\cdot\sigma(t)\,dt. \tag{95}$$

The following lemma provides an explicit expression for B_I.

Lemma 21. *For any* $x \in (-1,1)$,

$$B_I(\sigma)(x) = \pi^2\cdot\sigma(x) + \int_{-1}^{1}\frac{\log\frac{1-x}{1-t} - \log\frac{1+x}{1+t}}{x-t}\cdot\sigma(t)\,dt. \tag{96}$$

Proof. Substituting (90) into (95), we obtain

$$B_I(\sigma)(x) = \lim_{y\to 0}\int_{-1}^{1}\frac{|y|\cdot\left(\arctan\left(\frac{1-x}{|y|}\right)+\arctan\left(\frac{1+x}{|y|}\right)\right)}{((x-t)^2+y^2)}\cdot\sigma(t)\,dt$$
$$+\lim_{y\to 0}\int_{-1}^{1}\frac{(x-t)\cdot\left(\log\frac{(1-x)^2+y^2}{(1-t)^2} - \log\frac{(1+x)^2+y^2}{(1+t)^2}\right)}{2((x-t)^2+y^2)}\cdot\sigma(t)\,dt. \tag{97}$$

Combining (21) with the trivial identity

$$\lim_{y\to 0} \arctan\left(\frac{1-x}{|y|}\right) + \arctan\left(\frac{1+x}{|y|}\right) = \pi, \quad x \in (-1,1), \tag{98}$$

we have

$$\lim_{y\to 0} \int_{-1}^{1} \frac{|y| \cdot \left(\arctan\left(\frac{1-x}{|y|}\right) + \arctan\left(\frac{1+x}{|y|}\right)\right)}{((x-t)^2+y^2)} \cdot \sigma(t)\,dt = \pi^2 \cdot \sigma(x). \tag{99}$$

Now, applying Lebesgue's dominated convergence theorem (see, for example [18]) to the second part of the right-hand side of (97), we have

$$\begin{aligned} &\lim_{y\to 0} \int_{-1}^{1} \frac{(x-t) \cdot \left(\log\frac{(1-x)^2+y^2}{(1-t)^2} - \log\frac{(1+x)^2+y^2}{(1+t)^2}\right)}{2((x-t)^2+y^2)} \cdot \sigma(t)\,dt \\ &= \int_{-1}^{1} \lim_{y\to 0} \frac{(x-t) \cdot \left(\log\frac{(1-x)^2+y^2}{(1-t)^2} - \log\frac{(1+x)^2+y^2}{(1+t)^2}\right)}{2((x-t)^2+y^2)} \cdot \sigma(t)\,dt = \int_{-1}^{1} \frac{\log\frac{1-x}{1-t} - \log\frac{1+x}{1+t}}{x-t} \cdot \sigma(t)\,dt. \end{aligned} \tag{100}$$

Finally, combining (99), (100) with (97), we obtain (96). □

Remark 22. Elementary analysis shows that

$$\lim_{t\to x} \frac{\log\frac{1-x}{1-t} - \log\frac{1+x}{1+t}}{x-t} = -\frac{1}{1-x} - \frac{1}{1+x} = -\frac{2}{1-x^2}. \tag{101}$$

That is, the only singularities of the integral kernel in (96) are at the end points ± 1.

5.3. Connection between the operator B_I and the finite Hilbert transform

Lemma 23. *For any $\sigma \in F_I$ (see Definition 19),*

$$B_I(\sigma)(x) = -\widetilde{H}^2(\sigma)(x) \tag{102}$$

for all $x \in (-1,1)$.

Proof. Due to (12),

$$\widetilde{H}^2(\sigma)(x) = \int_{-1}^{1} \frac{1}{s-x} \cdot \left(\int_{-1}^{1} \frac{1}{t-s} \cdot \sigma(t)\,dt\right) ds. \tag{103}$$

Combining (37) with (103), we have

$$\widetilde{H}^2(\sigma)(x) = -\left(\pi^2 \cdot \sigma(x) + \int_{-1}^{1} \left(\int_{-1}^{1} \frac{1}{(s-x)(s-t)}\,ds\right) \cdot \sigma(t)\,dt\right). \tag{104}$$

Now, (102) follows immediately from the combination of (25), (96), (104). □

5.4. The inverse of $\widetilde{H}^2$ for Chebyshev polynomials

In Section 5.5, we will need the ability to solve equations of the form (19). However, due to Corollary 4, the solution to (19) is not unique. The purpose of this section is Theorem 28, stating that the solution to (19) *is* unique if restricted to the function space F_I (see Definition 19), and constructing such a solution.

The following lemma is a direct consequence of Corollary 4 and Lemma 6.

Lemma 24. *For any integer $n \geqslant 0$ and $x \in (-1,1)$, all solutions of the equation*

$$\widetilde{H}^2(\sigma_n) = T_n \tag{105}$$

are given by the formula

$$\sigma_n(x) = \widetilde{\sigma}_n(x) + \frac{C_0}{\sqrt{1-x^2}} + \frac{C_1}{\sqrt{1-x^2}} \cdot \log\frac{1+x}{1-x}, \tag{106}$$

with C_0, C_1 arbitrary constants, and the functions $\widetilde{\sigma}_n$ defined by the formulae:

$$\widetilde{\sigma}_0(x) = \frac{1}{\pi^3} \cdot \frac{x}{\sqrt{1-x^2}} \cdot \log\frac{1+x}{1-x}, \tag{107}$$

and

$$\widetilde{\sigma}_{2k}(x) = \frac{1}{\pi^3} \cdot \sqrt{1-x^2} \cdot \int_{-1}^{1} \frac{U_{2k-1}(t)}{t-x}\,\mathrm{d}t, \tag{108}$$

$$\widetilde{\sigma}_{2k-1}(x) = \frac{1}{\pi^3} \cdot \sqrt{1-x^2} \cdot \int_{-1}^{1} \frac{U_{2k-2}(t)}{t-x}\,\mathrm{d}t - \frac{2}{(2k-1)\pi^3} \cdot \frac{x}{\sqrt{1-x^2}}, \tag{109}$$

for all $k \geqslant 1$.

We will define the operators $J, L : C^1[-1,1] \to C[-1,1]$ via the formulae:

$$J(\varphi)(x) = \int_{-1}^{1} \log|x-t| \cdot \frac{1}{\sqrt{1-t^2}} \cdot \left(\int_{-1}^{1} \frac{\varphi(s)}{t-s}\,\mathrm{d}s \right) \mathrm{d}t, \tag{110}$$

$$L(\varphi)(x) = \int_{-1}^{1} \log|x-t| \cdot \sqrt{1-t^2} \cdot \left(\int_{-1}^{1} \frac{\varphi(s)}{t-s}\,\mathrm{d}s \right) \mathrm{d}t. \tag{111}$$

The following lemma provides explicit expressions for the derivatives of $J(\varphi)$, $L(\varphi)$, and for the values of $J(\varphi)$, $L(\varphi)$ at the points -1, 1.

Lemma 25. *For any $\varphi \in C^1[-1,1]$,*

$$J'(\varphi)(x) = -\pi^2 \cdot \frac{\varphi(x)}{\sqrt{1-x^2}}, \tag{112}$$

$$L'(\varphi)(x) = -\pi^2 \cdot \varphi(x) \cdot \sqrt{1-x^2} + \pi \cdot \int_{-1}^{1} \varphi(s)\,\mathrm{d}s, \tag{113}$$

for any $x \in (-1,1)$, and

$$J(\varphi)(-1) = \pi \cdot \int_{-1}^{1} \frac{\arccos(s)}{\sqrt{1-s^2}} \cdot \varphi(s)\,\mathrm{d}s, \tag{114}$$

$$J(\varphi)(1) = \pi \cdot \int_{-1}^{1} \frac{\arccos(s) - \pi}{\sqrt{1-s^2}} \cdot \varphi(s)\,\mathrm{d}s, \tag{115}$$

$$L(\varphi)(-1) = \pi \cdot \int_{-1}^{1} \varphi(x) \cdot (\arccos(x) \cdot \sqrt{1-x^2} + \log(2) \cdot x - 1)\,\mathrm{d}x. \tag{116}$$

$$L(\varphi)(1) = \pi \cdot \int_{-1}^{1} \varphi(x) \cdot ((\arccos(x) - \pi) \cdot \sqrt{1-x^2} + \log(2) \cdot x + 1)\,\mathrm{d}x. \tag{117}$$

Proof. The identities (114)–(117) are a direct consequence of (29)–(32) in Lemma 7, respectively. In order to prove (112), substituting (110) into $J'(\varphi)$ and interchanging the order of the differentiation and integration, we obtain

$$J'(\varphi)(x) = \int_{-1}^{1} \frac{1}{x-t} \cdot \frac{1}{\sqrt{1-t^2}} \cdot \left(\int_{-1}^{1} \frac{\varphi(s)}{t-s}\,\mathrm{d}s \right) \mathrm{d}t. \tag{118}$$

Applying (37) to the right-hand side of (118), we have

$$\begin{aligned} J'(\varphi)(x) = & -\pi^2 \cdot \frac{\varphi(x)}{\sqrt{1-x^2}} + \int_{-1}^{1} \frac{\varphi(s)}{x-s} \cdot \left(\int_{-1}^{1} \frac{1}{t-s} \cdot \frac{1}{\sqrt{1-t^2}}\,\mathrm{d}t \right) \mathrm{d}s \\ & - \int_{-1}^{1} \frac{\varphi(s)}{x-s} \cdot \left(\int_{-1}^{1} \frac{1}{t-x} \cdot \frac{1}{\sqrt{1-t^2}}\,\mathrm{d}t \right) \mathrm{d}s. \end{aligned} \tag{119}$$

Now, (112) follows immediately from the combination of (22), (119). The proof of (113) is virtually identical to that of (112). □

The following lemma provides explicit expressions for $J(T_n)$, with $n = 0, 1, 2, \ldots$

Lemma 26. *For any* $x \in [-1, 1]$,

$$J(T_0)(x) = -\frac{\pi^3}{2} + \pi^2 \cdot \arccos(x) \tag{120}$$

and

$$J(T_{2n})(x) = \frac{\pi^2}{2n} \cdot \sqrt{1-x^2} \cdot U_{2n-1}(x), \tag{121}$$

$$J(T_{2n-1})(x) = -\frac{2\pi}{(2n-1)^2} + \frac{\pi^2}{2n-1} \cdot \sqrt{1-x^2} \cdot U_{2n-2}(x) \tag{122}$$

for all $n \geqslant 1$.

Proof. Substituting T_0 into the Eqs. (112) and (114), we obtain

$$J'(T_0)(t)\,\mathrm{d}t = \frac{-\pi^2}{\sqrt{1-t^2}}, \tag{123}$$

$$J(T_0)(-1) = \pi \cdot \int_{-1}^{1} \frac{\arccos(s)}{\sqrt{1-s^2}}\,\mathrm{d}s = \pi \cdot \int_{0}^{\pi} x\,\mathrm{d}x = \frac{\pi^3}{2}. \tag{124}$$

Now, (120) follows immediately from the combination of (123) and (124), and the trivial identity

$$J(T_0)(x) = J(T_0)(-1) + \int_{-1}^{x} J'(T_0)(t)\,dt. \tag{125}$$

The proofs of (121), (122) are virtually identical to the proof of (120). □

The following lemma provides explicit expressions for $L(U_n)$, with $n = 0, 1, 2, \ldots$ It is a direct analogue of Lemma 26, replacing the mapping J with the mapping L, and the polynomials T_n with the polynomials U_n. Its proof is virtually identical to that of Lemma 26.

Lemma 27. *For any* $x \in [-1, 1]$,

$$L(U_0)(x) = \frac{\pi^2}{2} \cdot (\arccos x - x \cdot \sqrt{1-x^2}) + 2\pi \cdot x - \frac{\pi^3}{4} \tag{126}$$

and

$$L(U_{2n})(x) = \frac{\pi^2}{2} \cdot \sqrt{1-x^2} \cdot \left(\frac{U_{2n-1}(x)}{2n} - \frac{U_{2n+1}(x)}{2n+2} \right) + \frac{2\pi}{2n+1} \cdot x, \tag{127}$$

$$L(U_{2n-1})(x) = \frac{\pi^2}{2} \cdot \sqrt{1-x^2} \cdot \left(\frac{U_{2n-2}(x)}{2n-1} - \frac{U_{2n}(x)}{2n+1} \right) + 2\pi \cdot \left(\frac{2n \log 2}{4n^2-1} - \frac{4n}{(4n^2-1)^2} \right) \tag{128}$$

for all $n \geqslant 1$.

We are now in a position to combine the identities (27) and (28), Lemmas 24, 26 and 27 to obtain a refined version of Lemma 24. The following theorem is one of principal analytical tools of this paper.

Theorem 28. *Suppose that for each* $n = 0, 1, 2, \ldots$, *the function* $\sigma_n \in F_I$ (*see Definition* 19) *is the solution of the equation*

$$\tilde{H}^2(\sigma_n) = T_n. \tag{129}$$

Then

$$\sigma_0(x) = \frac{1}{\pi^3} \cdot \frac{x}{\sqrt{1-x^2}} \cdot \log \frac{1+x}{1-x} - \frac{2(\log 2 + 1)}{\pi^3 \log 2} \cdot \frac{1}{\sqrt{1-x^2}}, \tag{130}$$

$$\sigma_1(x) = \frac{1}{\pi^3} \cdot \sqrt{1-x^2} \cdot \int_{-1}^{1} \frac{U_0(t)}{t-x}\,dt - \frac{2}{\pi^3} \cdot \frac{x}{\sqrt{1-x^2}} + \frac{1}{2\pi^3} \cdot \frac{1}{\sqrt{1-x^2}} \cdot \log \frac{1+x}{1-x} \tag{131}$$

and

$$\sigma_{2n}(x) = \frac{1}{\pi^3} \cdot \sqrt{1-x^2} \cdot \int_{-1}^{1} \frac{U_{2n-1}(t)}{t-x}\,dt - \frac{2}{\pi^3 \log 2} \cdot \left(\frac{2n \log 2}{4n^2-1} - \frac{4n}{(4n^2-1)^2} \right) \cdot \frac{1}{\sqrt{1-x^2}}, \tag{132}$$

$$\sigma_{2n+1}(x) = \frac{1}{\pi^3} \cdot \sqrt{1-x^2} \cdot \int_{-1}^{1} \frac{U_{2n}(t)}{t-x}\,dt - \frac{2}{(2n+1)\pi^3} \cdot \frac{x}{\sqrt{1-x^2}} \tag{133}$$

for all $n \geqslant 1$.

Finally, we will need the following technical lemma.

Lemma 29. *Suppose that the functions $D_n : [-1,1] \to \mathbb{R}$ with $n = 0,1,2,\ldots$ are defined by the formula*

$$D_n(x) = \int_{-1}^{1} \log|x-t| \cdot \sigma_n(t)\,dt, \tag{134}$$

with σ_n defined by (130)–(133) *above.*

Then

$$D_0(x) = \frac{1}{\pi} \cdot \sqrt{1-x^2}, \tag{135}$$

$$D_1(x) = \frac{1}{2\pi} \cdot x \cdot \sqrt{1-x^2} \tag{136}$$

and

$$D_n(x) = \frac{1}{2\pi} \cdot \sqrt{1-x^2} \cdot \left(\frac{U_n(x)}{n+1} - \frac{U_{n-2}(x)}{n-1} \right), \tag{137}$$

for all $n \geqslant 2$.

Furthermore, for any integer $n \geqslant 2$, there exists a polynomial $p_{n-2}(x)$ of degree $n-2$ such that

$$D_n(x) = (1-x^2)^{3/2} \cdot p_{n-2}(x). \tag{138}$$

Proof. The identities (135)–(137) are a direct consequence of the identities (27) and (28), and Lemmas 26 and 27. To prove (138), we first observe that (see, for example [2]) for all $n = 0,1,2,\ldots,$

$$U_n(1) = n+1, \tag{139}$$

$$U_n(-1) = (-1)^n(n+1). \tag{140}$$

It immediately follows from 139 and 140 that

$$\frac{U_n(-1)}{n+1} - \frac{U_{n-2}(-1)}{n-1} = 0, \tag{141}$$

$$\frac{U_n(1)}{n+1} - \frac{U_{n-2}(1)}{n-1} = 0 \tag{142}$$

for any $n \geqslant 2$.

Now, we observe that the function

$$W(x) = \frac{U_n(x)}{n+1} - \frac{U_{n-2}(x)}{n-1} \tag{143}$$

is a polynomial of degree n, and that the points $x = \pm 1$ are the roots of W (see (141) and (142)). Therefore, there exists such a polynomial p_{n-2} of degree $n-2$ that

$$\frac{U_n(x)}{n+1} - \frac{U_{n-2}(x)}{n-1} = (1-x^2) \cdot p_{n-2}(x). \tag{144}$$

Finally, we obtain (138) by substituting (144) into (137). □

5.5. The integral equation formulation for the case of a line segment

In this section, we will combine the results in previous four sections to solve the Dirichlet problem for the line segment $I = [-1, 1]$ on the real axis. The following lemma is a direct consequence of Theorems 13 and 28, and Lemmas 23 and 29.

Lemma 30. *For any function $f \in C^2[-1,1]$, there exists a unique solution $\sigma \in F_I$ (see Definition 19) to the equation*

$$B_I(\sigma)(x) = \pi^2 \cdot \sigma(x) + \int_{-1}^{1} \frac{\log\frac{1-x}{1-t} - \log\frac{1+x}{1+t}}{x-t} \cdot \sigma(t)\,dt = f(x); \tag{145}$$

in other words, the operator B_I^{-1} is well defined if the range is restricted to the function space F_I. Furthermore, if f is orthogonal to T_0, T_1 with respect to the inner product (3), then the function $P_I(\sigma)$ can be continuously extended to $\mathbb{R}^2$.

For the cases $f = T_0$, $f = T_1$, we have the following lemma, easily verified by direct calculation.

Lemma 31.

1. *The only bounded continuous solution to the problem*

$$\begin{cases} \Delta u = 0 & \text{in } \mathbb{R}^2 \setminus I, \\ u = 1 & \text{on } I \end{cases} \tag{146}$$

is

$$u_I^0(x,y) = 1. \tag{147}$$

2. *The only bounded continuous solution to the problem*

$$\begin{cases} \Delta u = 0 & \text{in } \mathbb{R}^2 \setminus I, \\ u = x & \text{on } I \end{cases} \tag{148}$$

is

$$u_I^1(x,y) = \frac{N(x,y)}{D(x,y)}, \tag{149}$$

with the functions $N, D : \mathbb{R}^2 \to \mathbb{R}$ defined by the formulae

$$N(x,y) = \sqrt{(x+1)^2+y^2} - \sqrt{(x-1)^2+y^2}, \tag{150}$$

$$D(x,y) = \sqrt{(x+1)^2+y^2} + \sqrt{(x-1)^2+y^2} + \sqrt{\left(\sqrt{(x+1)^2+y^2} + \sqrt{(x-1)^2+y^2}\right)^2 - 4}, \tag{151}$$

respectively.

Combining Lemmas 30 and 31, we immediately obtain the following theorem.

Theorem 32. *Suppose that the function $f : [-1,1] \to \mathbb{R}$ is twice continuously differentiable. Suppose further that the function $\sigma \in F_I$ (see Definition 19), and the coefficients A_0, A_1 satisfy the following equations:*

$$B_I(\sigma)(x) = \pi^2 \cdot \sigma(x) + \int_{-1}^{1} \left\{ \left(\log\frac{1-x}{1-t} - \log\frac{1+x}{1+t} \right) \Big/ (x-t) \right\} \cdot \sigma(t)\,dt = f(x) - A_0 - A_1 \cdot x, \tag{152}$$

$$\int_{-1}^{1} (f(x) - A_0 - A_1 \cdot x) \cdot \frac{1}{\sqrt{1-x^2}}\,dx = 0, \tag{153}$$

$$\int_{-1}^{1} (f(x) - A_0 - A_1 \cdot x) \cdot \frac{x}{\sqrt{1-x^2}}\,dx = 0. \tag{154}$$

Then the function $u : \mathbb{R}^2 \to \mathbb{R}$ defined by the formula

$$u(x,y) = P_I(\sigma)(x,y) + A_0 \cdot u_I^0(x,y) + A_1 \cdot u_I^1(x,y) \tag{155}$$

is the solution of the problem

$$\begin{cases} \Delta u = 0 & \text{in } \mathbb{R}^2 \setminus I, \\ u = f & \text{on } I. \end{cases} \tag{156}$$

Applying Theorem 28, we can now solve the Dirichlet problem (156) via the representation (155).

Corollary 33. *Under the conditions of Theorem 32, the solutions to the Eqs. (152)–(154) are*

$$\sigma(x) = \frac{1}{\pi^3} \cdot \sqrt{1-x^2} \cdot \sum_{k=2}^{\infty} C_k \cdot \int_{-1}^{1} \frac{U_{k-1}(t)}{x-t}\,dt + \frac{B_0}{\sqrt{1-x^2}} + \frac{B_1 \cdot x}{\sqrt{1-x^2}}, \tag{157}$$

$$A_0 = C_0, \tag{158}$$

$$A_1 = C_1, \tag{159}$$

where the coefficients B_0, B_1 are defined by the formulae

$$B_0 = \frac{2}{\pi^3 \cdot \log 2} \cdot \sum_{k=1}^{\infty} C_{2k} \cdot \left(\frac{2k \log 2}{4k^2-1} - \frac{4k}{(4k^2-1)^2} \right), \tag{160}$$

$$B_1 = \frac{2}{\pi^3} \cdot \sum_{k=1}^{\infty} \frac{C_{2k+1}}{2k+1}, \tag{161}$$

respectively, and *C_k ($k = 0,1,2,\ldots$) are the Chebyshev coefficients of f given by* (9) *and* (10).

Remark 34. It immediately follows from Lemma 29 that the function $P_I(\sigma)$ with σ given by (157) has an explicit expression

$$P_I(\sigma)(x,y) = \int_{-1}^{1} \frac{(x-s)^2 - y^2}{((x-s)^2 + y^2)^2} \cdot D(\sigma)(s)\,ds, \tag{162}$$

for any $(x,y) \in \mathbb{R}^2 \setminus I$, with the function $D(\sigma) : [-1,1] \to \mathbb{R}$ defined by the formula

$$D(\sigma)(x) = \frac{1}{2\pi} \cdot \sqrt{1-x^2} \cdot \sum_{k=2}^{\infty} C_k \cdot \left(\frac{U_{k-2}(x)}{k-1} - \frac{U_k(x)}{k+1} \right). \tag{163}$$

Finally, we will need the following lemma.

Lemma 35. *Suppose that the operator S is defined by the formula*

$$S(\eta)(x) = D(B_I^{-1}(\eta))(x) = \int_{-1}^{1} \log|x-t| \cdot B_I^{-1}(\eta)(t)\,dt, \tag{164}$$

with the operator B_I defined in (96). *Then S is a bounded linear operator from $C[-1,1]$ to $C[-1,1]$.*

Proof. By Lemma 29, we have

$$S(T_0)(x) = -\frac{1}{\pi} \cdot \sqrt{1-x^2}, \tag{165}$$

$$S(T_1)(x) = -\frac{1}{2\pi} \cdot x \cdot \sqrt{1-x^2}, \tag{166}$$

and

$$S(T_n)(x) = -\frac{1}{2\pi} \cdot \sqrt{1-x^2} \cdot \left(\frac{U_n(x)}{n+1} - \frac{U_{n-2}(x)}{n-1} \right), \tag{167}$$

for all $n \geqslant 2$. Substituting (5) into (167), we obtain

$$S(T_n)(x) = -\frac{1}{2\pi} \cdot \left(\frac{\sin((n+1)\arccos(x))}{n+1} - \frac{\sin((n-1)\arccos(x))}{n-1} \right), \tag{168}$$

for all $n \geqslant 2$. Utilizing the trivial fact that $|\sin(u)| \leqslant 1$ for any real number u, we have

$$\|S(T_n)\|_\infty \leqslant \frac{2}{\pi} \cdot \frac{1}{n+1}, \tag{169}$$

for all $n = 0, 1, 2, \ldots$ Now, any function $\varphi \in C^2[-1,1]$ can be expanded into a Chebyshev series

$$\varphi(x) = \sum_{n=0}^{\infty} C_n \cdot T_n(x), \tag{170}$$

and by Parseval's identity,

$$\sum_{n=0}^{\infty} C_n^2 = \int_{-1}^{1} \frac{\varphi(x)^2}{\sqrt{1-x^2}}\,dx \leqslant \pi \cdot \|\varphi\|_\infty^2. \tag{171}$$

Applying Schwarz's inequality, we have

$$\|S(\varphi)\|_\infty \leqslant \sum_{n=0}^{\infty} |C_n| \cdot \|S(T_n)\|_\infty \leqslant \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{1}{n+1} \cdot |C_n| \leqslant \frac{2}{\pi} \left(\sum_{n=0}^{\infty} \frac{1}{(n+1)^2} \right)^{1/2} \cdot \left(\sum_{n=0}^{\infty} C_n^2 \right)^{1/2} \leqslant 2\|\varphi\|_\infty. \tag{172}$$

Since $C^2[-1,1]$ is dense in $C[-1,1]$, S is bounded from $C[-1,1]$ to $C[-1,1]$. □

6. Analytical apparatus III: open surface problem on a general curve

6.1. The integral operator P_γ

In this section, we consider the case of a general curve. We assume that $\gamma : [-1,1] \to \mathbb{R}^2$ is a sufficiently smooth "open" curve with the parametrization (71). The image of γ is denoted by Γ. We will consider the operator $P_\gamma : F_I \to C^2(\mathbb{R}^2 \setminus \Gamma)$ defined by the formula

$$\begin{aligned} P_\gamma(\sigma)(x) &= \int_{-1}^{1} K_\gamma(x,t)\cdot\sigma(t)\,\mathrm{d}t \\ &= \frac{L^2}{4}\cdot\int_{-1}^{1}\left(\frac{\partial^2 G_{\gamma(s)}(x)}{\partial N(s)^2} - c(s)\frac{\partial G_{\gamma(s)}(x)}{\partial N(s)}\right)\cdot\left(\int_{-1}^{1}\log|s-t|\sigma(t)\,\mathrm{d}t\right)\mathrm{d}s, \end{aligned} \tag{173}$$

with L the arc length of γ. The following lemma provides an explicit expression for the kernel K_γ. Its proof is virtually identical to that of Lemma 20.

Lemma 36. *For any $\sigma \in F_I$ (see Definition 19),*

$$K_\gamma(x,t) = \int_{-1}^{1}\frac{\partial G_{\gamma(s)}(x)}{\partial T(s)}\cdot\frac{1}{s-t}\,\mathrm{d}s, \tag{174}$$

for any $x \in \mathbb{R}^2 \setminus \Gamma$ and $t \in (-1,1)$, with the integral in (174) intepreted in the principal value sense.

6.2. The boundary integral operator B_γ

We will then define the integral operator $B_\gamma : F_I \to L^1[-1,1]$ by the formula

$$B_\gamma(\sigma)(t) = \lim_{h\to 0} P_\gamma(\sigma)(\gamma(t)+h\cdot N(t)) = \lim_{h\to 0}\int_{-1}^{1} K_\gamma(\gamma(t)+h\cdot N(t),s)\cdot\sigma(s)\,\mathrm{d}s. \tag{175}$$

The following lemma is a direct consequence of Lemmas 12 and 21; it provides an explicit expression for B_γ.

Lemma 37. *For any $t \in (-1,1)$,*

$$B_\gamma(\sigma)(t) = \pi^2\cdot\sigma(t) + \int_{-1}^{1} K_\gamma^b(t,s)\cdot\sigma(s)\,\mathrm{d}s, \tag{176}$$

with the kernel $K_\gamma^b : (-1,1)\times(-1,1) \to \mathbb{R}$ given by the formula

$$K_\gamma^b(t,s) = \int_{-1}^{1}\frac{\partial G_{\gamma(x)}(\gamma(t))}{\partial T(x)}\cdot\frac{1}{x-s}\,\mathrm{d}x, \tag{177}$$

with the integral in (177) intepreted in the principal value sense.

6.3. The integral equation formulation for the case of a general curve

Similarly to the operator B_I defined in (96), the kernel K_γ^b of B_γ is strongly singular at the end-points. Therefore, if the solution of the Dirichlet problem (69) and (70) is represented by the function $P_\gamma(\sigma)$ on $\mathbb{R}^2 \setminus \Gamma$, then (70) will lead to a boundary integral equation

$$B_\gamma(\sigma)(t) = f(t), \tag{178}$$

which is *not* of the second kind. Because of the obvious similarity of the operators B_I, B_γ, it is natural to consider the operator $\widetilde{P}_\gamma : C[-1,1] \to C^2(\mathbb{R}^2 \setminus \Gamma)$ defined by the formula

$$\widetilde{P}_\gamma(\eta)(x) = P_\gamma \circ B_I^{-1}(\eta)(x). \tag{179}$$

Obviously, $\widetilde{P}_\gamma(\eta)$ is not defined when $x \in \Gamma$, and we will define the operator $\widetilde{B}_\gamma : C[-1,1] \to C[-1,1]$ by the formula

$$\widetilde{B}_\gamma(\eta)(t) = \lim_{x \to \gamma(t)} \widetilde{P}_\gamma(\eta)(x) = B_\gamma \circ B_I^{-1}(\eta)(t). \tag{180}$$

The following theorem is one of principal results of the paper; it states that $\widetilde{B}_\gamma$ is a *second kind* integral operator when restricted to continuous functions, and is an immediate consequence of Lemmas 15 and 35.

Theorem 38. *Suppose that $\gamma : [-1,1] \to \mathbb{R}^2$ is a sufficiently smooth* "open" curve with the parametrization (71). *Suppose further that the operator $\widetilde{R}_\gamma : C[-1,1] \to C[-1,1]$ is defined by the formula*

$$\widetilde{R}_\gamma(\sigma)(x) = \int_{-1}^{1} \widetilde{r}(x,t) \cdot \sigma(t)\,dt, \tag{181}$$

with the function $\widetilde{r} : [-1,1] \times [-1,1] \to \mathbb{R}$ defined by the formula

$$\widetilde{r}(x,t) = \frac{L^2}{4} \cdot \left(-\frac{2\langle N(t), \gamma(x) - \gamma(t)\rangle^2}{\|\gamma(x) - \gamma(t)\|^4} + \frac{1}{\|\gamma(x) - \gamma(t)\|^2} \right) + \frac{L^2 \cdot c(t)}{4} \cdot \frac{\langle N(t), \gamma(x) - \gamma(t)\rangle}{\|\gamma(x) - \gamma(t)\|^2} - \frac{1}{(x-t)^2}, \tag{182}$$

for all $x \neq t$, and by the formula

$$\widetilde{r}(t,t) = \frac{L^2 \cdot c(t)^2}{48}, \tag{183}$$

for all $x = t$, with L the arc length of γ, and $c(t)$ the curvature of γ at the point $\gamma(t)$. Then,

$$\widetilde{B}_\gamma(\eta)(t) = (I + M)(\eta)(t), \tag{184}$$

with $I : C[-1,1] \to C[-1,1]$ the identity operator, and $M : C[-1,1] \to C[-1,1]$ a compact operator defined by the formula

$$M(\eta)(t) = (B_\gamma - B_I) \circ B_I^{-1}(\eta)(t) = \widetilde{R}_\gamma \circ S(\eta)(t), \tag{185}$$

with the operators $\widetilde{R}_\gamma, S : C[-1,1] \to C[-1,1]$ defined by ((181), (164)–(167)), *respectively.*

Remark 39. It immediately follows from the combination of (59) and (182) that the operator $\widetilde{R}_\gamma$ is related to R_γ defined in Lemma 15 by the formula

$$\widetilde{R}_\gamma(\widetilde{\sigma})(x) = \frac{L}{2} \cdot R_\gamma(\sigma)\left(\frac{L}{2}(x+1)\right), \tag{186}$$

with $\widetilde{\sigma}(t) = \sigma(\frac{L}{2}(t+1))$, and the function $\widetilde{r}$ is related to the function r defined in (59) by the formula

$$\widetilde{r}(x,t) = \frac{L^2}{4} \cdot r\left(\frac{L}{2}(x+1), \frac{L}{2}(t+1)\right). \tag{187}$$

The function $\widetilde{P}_\gamma(\eta)$ cannot, in general, be continuously extended to the whole plane $\mathbb{R}^2$, unless the density η satisfies certain additional conditions. The following lemma is a direct consequence of Theorems 13 and 28, and Lemmas 23 and 29.

Lemma 40. *Suppose that the function $\eta \in C[-1,1]$ is orthogonal to T_0 and T_1 with respect to the inner product (3). Then $\widetilde{P}_\gamma(\eta)$ can be continuously extended to $\mathbb{R}^2$.*

Lemma 40 above shows that the solution of the problem (69) and (70) cannot be represented by the function $\widetilde{P}_\gamma(\eta)$ alone. Indeed, $\widetilde{P}_\gamma(\eta)(x)$ decays at infinity like $1/|x|$, whereas Theorem 17 only requires that the solution of the problem (69) and (70) be bounded at infinity. Suppose now that we can find two functions u_γ^0, u_γ^1 in S_γ (see (68)) such that the following condition holds:

$$\det \begin{pmatrix} \langle \eta_0, T_0 \rangle & \langle \eta_0, T_1 \rangle \\ \langle \eta_1, T_0 \rangle & \langle \eta_1, T_1 \rangle \end{pmatrix} \neq 0, \tag{188}$$

with η_0, η_1 the solutions to the equations

$$\widetilde{B}_\gamma(\eta_0)(t) = u_\gamma^0(\gamma(t)), \tag{189}$$

$$\widetilde{B}_\gamma(\eta_1)(t) = u_\gamma^1(\gamma(t)), \tag{190}$$

respectively, and the inner product in (188) defined by (3). Then the solution of the problem (69) and (70) can be represented by the formula

$$u(x) = \widetilde{P}_\gamma(\eta)(x) + A_0 \cdot u_\gamma^0(x) + A_1 \cdot u_\gamma^1(x), \tag{191}$$

so that the density η, while satisfying the boundary integral equation

$$\widetilde{B}_\gamma(\eta)(t) = f(t) - A_0 \cdot u_\gamma^0(\gamma(t)) - A_1 \cdot u_\gamma^1(\gamma(t)), \tag{192}$$

is also orthogonal to T_0 and T_1. The following lemma provides such two functions indirectly; it describes a single-layer-potential representation for the functions $\widetilde{P}_\gamma(T_n)$ $(n = 2, 3, \ldots)$.

Lemma 41. *Suppose that $\gamma : [-1,1] \to \mathbb{R}^2$ is a sufficiently smooth* "open" curve with the parametrization (71). *Then for any $n = 2, 3, \ldots$,*

$$\widetilde{P}_\gamma(T_n)(x) = -\frac{n}{\pi} \cdot \int_{-1}^{1} G_{\gamma(t)}(x) \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\,\mathrm{d}t = -\frac{n}{\pi} \cdot \int_{-1}^{1} \log|x - \gamma(t)| \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\,\mathrm{d}t, \tag{193}$$

for any $x \notin \Gamma$.

Proof. Combining (179), (173), (164), we have the identity

$$\widetilde{P}_\gamma(\eta)(x) = \frac{L^2}{4} \cdot \int_{-1}^{1} \left(\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2} - c(t) \frac{\partial G_{\gamma(t)}(x)}{\partial N(t)} \right) \cdot S(\eta)(t)\,\mathrm{d}t, \tag{194}$$

for an arbitrary $\eta \in C[-1,1]$. In particular,

$$\widetilde{P}_\gamma(T_n)(x) = \frac{L^2}{4} \cdot \int_{-1}^{1} \left(\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2} - c(t) \frac{\partial G_{\gamma(t)}(x)}{\partial N(t)} \right) \cdot S(T_n)(t)\,\mathrm{d}t. \tag{195}$$

Since the function $G_{\gamma(t)}(x)$ satisfies the Laplace equation for all $x \neq \gamma(t)$, applying (33) to $G_{\gamma(t)}$ and carrying out elementary analytic manipulations, we obtain the identity

$$\frac{L^2}{4} \cdot \left(\frac{\partial^2 G_{\gamma(t)}(x)}{\partial N(t)^2} - c(t) \frac{\partial G_{\gamma(t)}(x)}{\partial N(t)} \right) = - \frac{\partial^2 G_{\gamma(t)(x)}}{\partial t^2}, \tag{196}$$

and substitution of (196), (167) into (195) yields the identity

$$\widetilde{P}_\gamma(T_n)(x) = \frac{1}{2\pi} \int_{-1}^{1} \frac{\partial^2 G_{\gamma(t)(x)}}{\partial t^2} \cdot \sqrt{1-t^2} \cdot \left(\frac{U_n(t)}{n+1} - \frac{U_{n-2}(t)}{n-1} \right) \mathrm{d}t. \tag{197}$$

Now, we obtain (193) by integrating by parts twice the right-hand side of (197). □

The following lemma is an immediate consequence of Lemma 41 and the well-known fact that the functions $u_\gamma^n : \mathbb{R}^2 \to \mathbb{R}$ $(n = 0, 1, 2, \ldots)$ defined by the formulae

$$u_\gamma^0(x) = 1, \tag{198}$$

$$u_\gamma^n(x) = \int_{-1}^{1} \log|x - \gamma(t)| \cdot \frac{T_n(t)}{\sqrt{1-t^2}} \mathrm{d}t, \quad n = 1, 2, \ldots \tag{199}$$

form a complete basis for the space S_γ (see, for example [15]).

Lemma 42. *Suppose that* $\gamma : [-1, 1] \to \mathbb{R}^2$ *is a sufficiently smooth* "open" curve with the parametrization (71). *Then the functions* u_γ^0, u_γ^1 *defined by* (198) *and* (199) *satisfy the condition* (188)–(190).

Finally, we summarize our analysis for the case of a general curve by the following theorem.

Theorem 43. *Suppose that* $\gamma : [-1, 1] \to \mathbb{R}^2$ *is a sufficiently smooth* "open" curve with the parametrization (71), *and that the function* $f : [-1, 1] \to \mathbb{R}$ *is twice continuously differentiable. Suppose further that the function* $\eta : [-1, 1] \to \mathbb{R}$, *and the coefficients* A_0, A_1 *satisfy the equations*

$$\widetilde{B}_\gamma(\eta)(t) = (I + \widetilde{R}_\gamma \circ S)(\eta)(t) = f(t) - A_0 \cdot u_\gamma^0(\gamma(t)) - A_1 \cdot u_\gamma^1(\gamma(t)), \tag{200}$$

$$\int_{-1}^{1} \eta(t) \cdot \frac{1}{\sqrt{1-t^2}} \mathrm{d}t = 0, \tag{201}$$

$$\int_{-1}^{1} \eta(t) \cdot \frac{t}{\sqrt{1-t^2}} \mathrm{d}t = 0, \tag{202}$$

with I *the identity operator, and the operators* $\widetilde{R}_\gamma, S : C[-1, 1] \to C[-1, 1]$ *defined by* (181) *and* (164), *respectively. Then the function* $u : \mathbb{R}^2 \to \mathbb{R}$ *defined by the formula*

$$u(x) = \widetilde{P}_\gamma(\eta)(x) + A_0 \cdot u_\gamma^0(x) + A_1 \cdot u_\gamma^1(x) \tag{203}$$

is the solution of the problem

$$\begin{cases} \Delta u = 0 & \text{in } \mathbb{R}^2 \setminus \Gamma, \\ u = f & \text{on } \Gamma, \end{cases} \tag{204}$$

in (203), *the operator* $\widetilde{P}_\gamma : C[-1, 1] \to C(\mathbb{R}^2)$ *is defined by* (179), (173), (145), *and the functions* u_γ^0, u_γ^1 *are defined by* (198) *and* (199) *respectively.*

Remark 44. For the case of several open curves $\Gamma = \sum_{i=1}^{m} \gamma_i$, the following modifications should be made. Instead of (203), the function u will be given by the formula

$$u(x) = \sum_{i=1}^{m} \left\{ \widetilde{P}_{\gamma_i}(\eta_i)(x) + A_0^i \cdot u_{\gamma_i}^0(x) + A_1^i \cdot u_{\gamma_i}^1(x) \right\} + C, \tag{205}$$

with C a real number to be determined, and the functions $u_{\gamma_i}^n$ $(i = 1, \ldots, m;\ n = 0, 1)$ defined by the formula

$$u_{\gamma_i}^n(x) = \int_{-1}^{1} \log|x - \gamma_i(t)| \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\, dt. \tag{206}$$

The functions η_i, and the coefficients A_0^i, A_1^i, C are determined as the solution of the system of equations

$$\widetilde{B}_{\gamma_i}(\eta_i)(t) = (I + \widetilde{R}_{\gamma_i} \circ S)(\eta_i)(t) = f_i(t) - \sum_{j=1}^{m} \left\{ A_0^j \cdot u_{\gamma_j}^0(\gamma_i(t)) - A_1^j \cdot u_{\gamma_j}^1(\gamma_i(t)) \right\} - C, \tag{207}$$

$$\int_{-1}^{1} \eta_i(t) \cdot \frac{1}{\sqrt{1-t^2}}\, dt = 0, \tag{208}$$

$$\int_{-1}^{1} \eta_i(t) \cdot \frac{t}{\sqrt{1-t^2}}\, dt = 0, \tag{209}$$

$$\sum_{i=1}^{m} A_0^i = 0. \tag{210}$$

Clearly, the functions $u_{\gamma_i}^0$ defined by (206) are linearly independent; the constraint (210) and the constant term C are introduced so that the function u is bounded at infinity.

7. Numerical algorithm

In this section, we construct a rudimentary numerical algorithm for the solution of the Dirichlet problem (69) and (70) via the Eqs. (200)–(202). Since the construction of the matrix and the solver of the resulting linear system are direct, the algorithm requires $O(N^3)$ work and $O(N^2)$ storage, with N the number of nodes on the boundary. While standard acceleration techniques (such as the Fast Multipole Method, etc.) could be used to improve these estimates, no such acceleration was performed, since the purpose of this section (as well as the following one) is to demonstrate the stability of the integral formulation and the convergence rate of a very simple discretization scheme.

By Theorem 43, the equations to be solved are (200)–(202), where the unknowns are the function η, and two real numbers A_0, A_1. To solve (200)–(202) numerically, we discretize the boundary into N Chebyshev nodes and approximate the unknown density η by a finite Chebyshev series of the first kind,

$$\eta(t) \simeq \sum_{k=0}^{N-1} C_k \cdot T_k(t), \tag{211}$$

with the coefficients C_k $(k = 0, \ldots, N-1)$ to be determined. In order to discretize (200), we start with observing that by (165)–(167), the action of the operator S on the function η is described via the formula

$$S(\eta)(x) = \sum_{k=0}^{N-1} \left(\sum_{j=0}^{N-1} B_{kj} \cdot C_j \right) \cdot \frac{2}{\pi} \cdot U_k(x) \cdot \sqrt{1-x^2}, \tag{212}$$

where the matrix $B = (B_{kj})$ $(k, j = 0, \ldots, N-1)$ is given by the formulae

$$\begin{cases} B_{00} = -\frac{1}{2}, \\ B_{kk} = -\frac{1}{4k}, & 1 \leqslant k \leqslant N-1, \\ B_{k,k+2} = \frac{1}{4k}, & 0 \leqslant k \leqslant N-3, \\ B_{kj} = 0, & \text{otherwise.} \end{cases} \tag{213}$$

In other words, given a function η expressed as a Chebyshev series of the first kind, (212) expresses $S(\eta)$ as a Chebyshev series of the second kind. Now, it is natural to approximate the operator $\widetilde{R}_\gamma$ by an expression converting functions of the form

$$\sum_{k=0}^{N-1} \alpha_k \cdot U_k(t) \tag{214}$$

into functions of the form

$$\sum_{k=0}^{N-1} \beta_k \cdot T_k(x), \tag{215}$$

with the product $\widetilde{R}_\gamma \circ S$ converting expressions of the form (215) into expressions of the same form. Thus, we approximate the kernel $\widetilde{r}(x,t)$ (see (182)) of the operator $\widetilde{R}_\gamma$ with an expression of the form

$$\widetilde{r}(x,t) \simeq \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} K_{ij} \cdot T_i(x) \cdot U_j(t). \tag{216}$$

Clearly, the coefficients K_{ij} have to be determined numerically, since the curve Γ is user-specified, and is unlikely to have a convenient analytical expression. Thus, we obtain the coefficients K_{ij} by first constructing the $N \times N$ matrix $R = (\widetilde{r}(x_i, t_j))$ $(i, j = 0, 1, \ldots, N-1)$ with $x_i (i = 0, 1, \ldots, N-1)$ the Chebyshev nodes defined by (4) and $t_j (j = 0, \ldots, N-1)$ the Chebyshev nodes of the second kind defined by (7), then converting R into the matrix $K = (K_{ij})(i, j = 0, 1, \ldots, N-1)$ by the formula

$$K = U \cdot R \cdot V, \tag{217}$$

with $N \times N$ matrices $U = (U_{ij})$, $V = (V_{ij})$ defined by the formulae

$$\begin{cases} U_{0j} = \frac{1}{N} \cdot T_0(x_j), & j = 0, 1, \ldots, N-1, \\ U_{ij} = \frac{2}{N} \cdot T_i(x_j), & i = 1, \ldots, N-1, \quad j = 0, 1, \ldots, N-1, \end{cases} \tag{218}$$

$$V_{ij} = \frac{2}{N+1} \cdot \sin^2 \left(\frac{(N-i) \cdot \pi}{N+1} \right) \cdot U_j(t_i), \quad i, j = 0, 1, \ldots, N-1, \tag{219}$$

respectively. We then approximate the prescribed Dirichlet data f by its Chebyshev approximation of order $N-1$

$$f(t) \simeq \sum_{k=0}^{N-1} \hat{f}_k \cdot T_k(t), \tag{220}$$

where the coefficients $\hat{f}_k$ can be obtained by first evaluating f at Chebyshev nodes x_i, then applying to it the matrix U defined by (218), i.e.,

$$\hat{f}_k = \sum_{i=0}^{N-1} U_{ki} \cdot f(x_i). \tag{221}$$

Similarly, we approximate the function u_γ^1 (see (199)) with an expression of the form

$$u_\gamma^1(\gamma(t)) \simeq \sum_{k=0}^{N-1} \hat{u}_k \cdot T_k(t), \tag{222}$$

with the coefficients $\hat{u}_k$ defined by the formula

$$\hat{u}_k = \sum_{i=0}^{N-1} U_{ki} \cdot u_\gamma^1(\gamma(x_i)), \tag{223}$$

with x_i the Chebyshev nodes defined by (4). Combining (212), (216), (221), and (222), we discretize (200) into the equation

$$\widetilde{A} \cdot \begin{pmatrix} C_0 \\ C_1 \\ \vdots \\ C_{N-1} \end{pmatrix} + A_0 \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + A_1 \cdot \begin{pmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \vdots \\ \hat{u}_{N-1} \end{pmatrix} = \begin{pmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \vdots \\ \hat{f}_{N-1} \end{pmatrix}, \tag{224}$$

with $N \times N$ matrix $\widetilde{A}$ defined by the formula

$$\widetilde{A} = I_N + K \cdot B, \tag{225}$$

with I_N the $N \times N$ identity matrix. Furthermore, (201) and (202) lead to the equations

$$C_0 = 0, \tag{226}$$

$$C_1 = 0. \tag{227}$$

Finally, combining (224), (226), (227), we obtain the following linear system of dimension $N+2$ to be solved

$$\begin{pmatrix} 1\,0\,0\ldots0 & 0 & 0 \\ 0\,1\,0\ldots0 & 0 & 0 \\ & 1 & \hat{u}_0 \\ \widetilde{A} & 0 & \hat{u}_1 \\ & \vdots & \vdots \\ & 0 & \hat{u}_{N-1} \end{pmatrix} \cdot \begin{pmatrix} C_0 \\ C_1 \\ \vdots \\ C_{N-1} \\ A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \hat{f}_0 \\ \hat{f}_1 \\ \vdots \\ \hat{f}_{N-1} \end{pmatrix}. \tag{228}$$

Remark 45. Having solved (228) with any standard solver (we used DGECO from LINPACK), we can compute the solution of the Problem (69) and (70) at any point in $\mathbb{R}^2$ via (203).

Remark 46. The algorithm can be generalized to the case when the boundary consists of several disjoint open curves, and the generalization is straightforward (see Remark 44).

8. Numerical examples

A FORTRAN code has been written implementing the algorithm described in the preceding section. In this section, we demonstrate the performance of the scheme with several numerical examples. We consider the problem in electrostatics: the boundary is made of conductor and grounded, the electric field incident on the boundary is generated by the sources outside the boundary; that is to say, there are three fields present: the incident field u_i, the reflected field u_r, and the total field $u_t = u_i + u_r$, where $u_t = 0$ on the boundary, and $u_r = -u_i$ on the boundary and is harmonic elsewhere. For these examples, we plot the equipotential lines of the total field and present tables showing the convergence rate of the algorithm by computing the errors of the reflected field.

Remark 47. In the examples below, the problems to be solved via the procedure of the preceding section have no simple analytical solution. Thus, we tested the accuracy of our procedure by evaluating our solution via the formula (203) at a large number M of nodes on the boundary Γ (in our experiments, we always used $M = 4000$), and comparing it with the analytically evaluated right-hand side. We did not need to verify the fact that our solutions satisfy the Laplace equation, since this follows directly from the representation (203).

In each of those tables, the first column contains the total number N of nodes in the discretization of each curve. The second column contains the condition number of the linear system. The third column contains the relative L^2 error of the numerical solution as compared with the analytically evaluated Dirichlet data on the boundary. The fourth column contains the maximum absolute error on the boundary. In the last two columns, we list the errors of the numerical solution as compared with the numerical solution with twice the number of nodes, where the solution is evaluated at 4000 equispaced points on a circle of radius 1.4 centered at the origin; the fifth column contains the relative L^2 error, and the sixth column contains the maximum absolute error.

Example 1. In this example, the boundary is the line segment parametrized by the formula

$$\begin{cases} x(t) = t, \\ y(t) = -0.2, \end{cases} \quad -1 \leqslant t \leqslant 1. \tag{229}$$

The Dirichlet data are generated by a unit charge at (0,0). The numerical results are shown in Table 1. The source, curve and equipotential lines are plotted in Fig. 1.

Example 2. In this example, the boundary is an elliptic arc parametrized by the formula

$$\begin{cases} x(t) = 0.8\cos(t), \\ y(t) = 0.5\sin(t) + 0.25, \end{cases} \quad -\pi \leqslant t \leqslant 0. \tag{230}$$

Table 1
Numerical results for Example 1

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
4	$0.524E+01$	$0.288E+00$	$0.607E+00$	$0.513E-01$	$0.590E-01$
8	$0.450E+01$	$0.703E-01$	$0.178E+00$	$0.613E-02$	$0.686E-02$
16	$0.388E+01$	$0.759E-02$	$0.212E-01$	$0.133E-03$	$0.146E-03$
32	$0.344E+01$	$0.165E-03$	$0.486E-03$	$0.115E-06$	$0.126E-06$
64	$0.318E+01$	$0.147E-06$	$0.446E-06$	$0.146E-12$	$0.164E-12$
128	$0.303E+01$	$0.252E-12$	$0.839E-12$	$0.250E-13$	$0.265E-13$

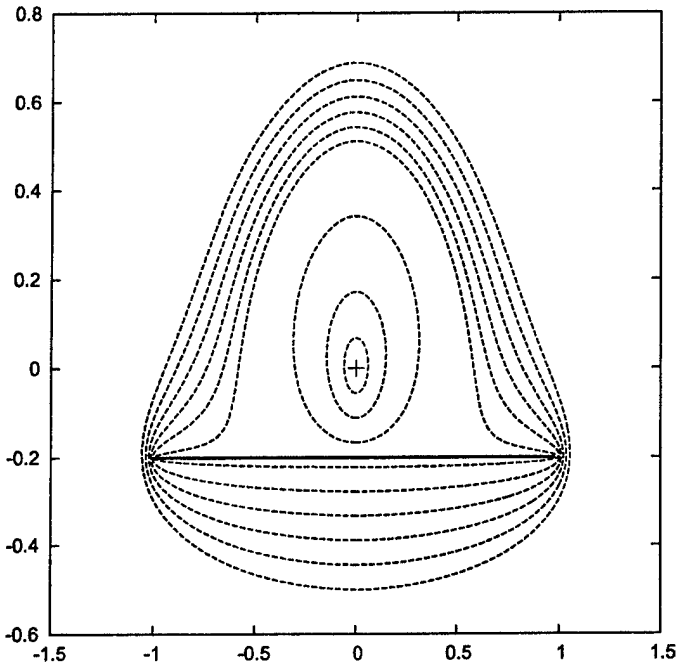


Fig. 1. Source, curve, and equipotential lines for Example 1.

Table 2
Numerical results for Example 2

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
4	$0.513E+01$	$0.180E+00$	$0.124E+00$	$0.343E-01$	$0.166E-01$
8	$0.461E+01$	$0.722E-01$	$0.554E-01$	$0.668E-02$	$0.333E-02$
16	$0.399E+01$	$0.103E-01$	$0.833E-02$	$0.155E-03$	$0.773E-04$
32	$0.352E+01$	$0.230E-03$	$0.187E-03$	$0.855E-07$	$0.426E-07$
64	$0.316E+01$	$0.128E-06$	$0.105E-06$	$0.475E-13$	$0.201E-13$
128	$0.301E+01$	$0.141E-12$	$0.134E-12$	$0.272E-13$	$0.102E-13$

The Dirichlet data are generated by one positive charge of unit strength at (0,0) and another negative charge of unit strength at (0,–0.5). The numerical results are shown in Table 2. The sources, curve, and equipotential lines are plotted in Fig. 2.

Example 3. In this example, the boundary is a spiral parametrized by the formula

$$\begin{cases} x(t) = t\cos(3.3t) - 0.1, \\ y(t) = t\sin(3.3t), \end{cases} \quad 0.2 \leqslant t \leqslant 1.2. \tag{231}$$

The Dirichlet data are generated by a unit charge at (0,0). The numerical results are shown in Table 3. The source, curve, and equipotential lines are plotted in Fig. 3.

Example 4. In this example, we consider the case of several open curves. The boundary consists of three elliptic arcs parametrized by the formulae

$$\begin{cases} x_1(t) = 1.1\cos(t) - 1, \\ y_1(t) = \sin(t) + 0.5, \end{cases} \quad -\frac{\pi}{12} \leqslant t \leqslant \frac{\pi}{4}, \tag{232}$$

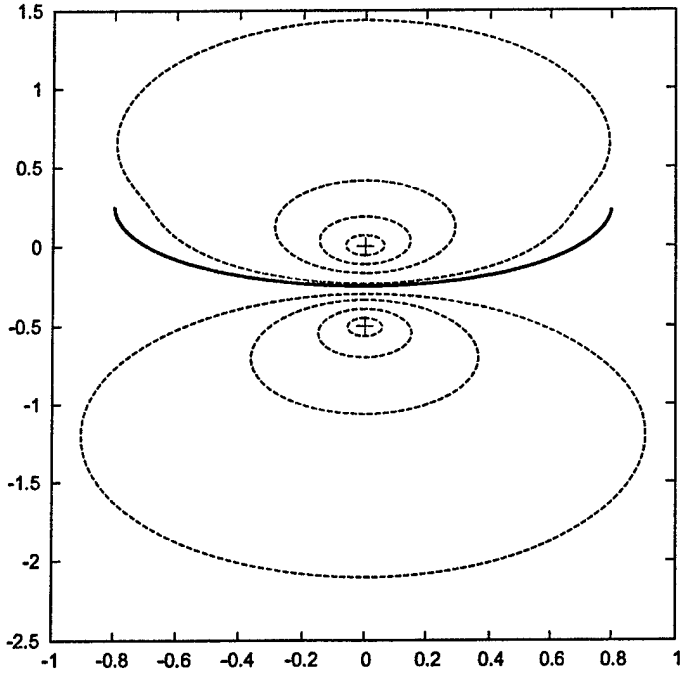


Fig. 2. Sources, curve, and equipotential lines for Example 2.

Table 3
Numerical results for Example 3

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
8	$0.325E+02$	$0.215E-01$	$0.323E-01$	$0.478E+00$	$0.426E+00$
16	$0.579E+01$	$0.549E-03$	$0.986E-03$	$0.658E-01$	$0.820E-01$
32	$0.478E+01$	$0.211E-05$	$0.317E-05$	$0.149E-02$	$0.194E-02$
64	$0.424E+01$	$0.987E-11$	$0.122E-10$	$0.350E-06$	$0.453E-06$
128	$0.392E+01$	$0.861E-13$	$0.520E-12$	$0.127E-12$	$0.119E-12$
256	$0.374E+01$	$0.138E-12$	$0.139E-11$	$0.139E-12$	$0.123E-12$

$$\begin{cases} x_2(t) = 1.1\cos(t), \\ y_2(t) = \sin(t) - 1.2, \end{cases} \quad \frac{7\pi}{12} \leqslant t \leqslant \frac{11\pi}{12}, \tag{233}$$

$$\begin{cases} x_3(t) = 1.1\cos(t) + 1, \\ y_3(t) = \sin(t) + 0.5, \end{cases} \quad -\frac{3\pi}{4} \leqslant t \leqslant -\frac{5\pi}{12}. \tag{234}$$

The Dirichlet data are generated by a unit charges at (0,0). The numerical results are shown in Table 4, where N is the number of nodes on each curve. The source, curves, and equipotential lines are plotted in Fig. 4.

Remark 48. The above examples illustrate the superalgebraic convergence of the scheme for smooth data and curves (see Remark 16 in Section 2.6). The number of nodes needed depends on the complexity of the underlying geometry and the smoothness of the prescribed data. The condition number of the resulting linear system is usually very low.

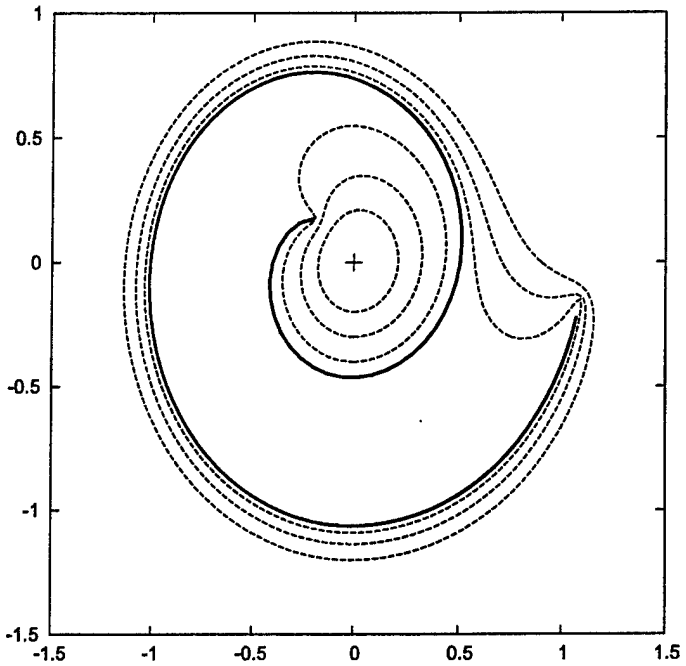


Fig. 3. Source, curve, and equipotential lines for Example 3.

Table 4
Numerical results for Example 4

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
4	$0.845E+01$	$0.113E-01$	$0.228E-01$	$0.493E-03$	$0.117E-02$
8	$0.754E+01$	$0.126E-03$	$0.269E-03$	$0.159E-05$	$0.108E-04$
16	$0.689E+01$	$0.173E-07$	$0.390E-07$	$0.656E-10$	$0.452E-09$
32	$0.649E+01$	$0.443E-12$	$0.196E-11$	$0.950E-13$	$0.113E-12$
64	$0.627E+01$	$0.658E-13$	$0.295E-12$	$0.492E-14$	$0.433E-14$
128	$0.615E+01$	$0.880E-13$	$0.356E-12$	$0.968E-14$	$0.971E-14$

9. Conclusions and generalizations

We have presented a stable second kind integral equation formulation for the Dirichlet problem for the Laplace equation in two dimensions, with the boundary condition specified on a curve (consisting of one or more separate segments). The resulting numerical algorithm converges superalgebraically if both the boundary data and the curves are smooth. Obviously, the combination of the Fast Multipole Method (see, for example, [7]) and any standard iterative solver yields an $O(N)$ algorithm, with N the number of nodes on the boundary.

The extensions of the scheme of this paper to other boundary conditions (such as Neumann condition, Robin condition, etc.) specified on an open curve Γ in $\mathbb{R}^2$ are fairly straightforward. For the Neumann problem, representing the solution in the form of a double layer potential, one obtains a hypersingular integral equation on Γ. Its subsequent preconditioning by a single layer potential yields a second kind integral equation (SKIE). For a Robin problem, one obtains an SKIE formulation by representing the

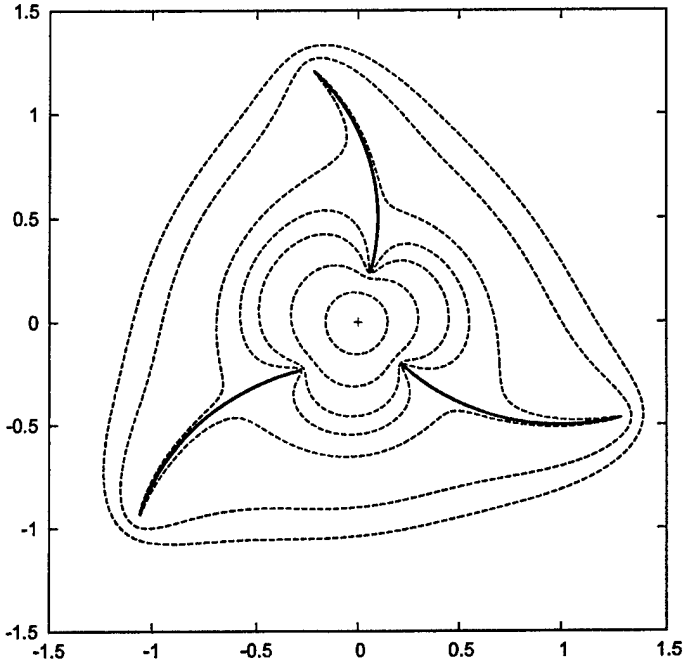


Fig. 4. Source, curves, and equipotential lines for Example 4.

solution via an appropriate linear combination of single and double layer potentials, with a further preconditioning by a single layer potential. Furthermore, the approach of this paper can be applied almost without modification to elliptic PDEs other than the Laplace equation (such as Helmholtz equation, Yukawa equaiton, etc.). Indeed, the Green's function for any such equation has the form

$$G(x,y) = \phi(x,y) \cdot \log(||x-y||) + \psi(x,y), \tag{235}$$

with ϕ, ψ a pair of smooth functions, and $\phi(0,0) = 1/(2\pi)$ (see, for example [4]). When the procedure of Section 6 of this paper is applied to a Green's function of the form (235), the result is virtually identical to that obtained in Section 6.3, except for the change in the compact operator $\widetilde{P}_\gamma$ in (179). However, the convergence rate of the numerical scheme of Section 7 deteriorates drastically, since in this case the kernel K of the operator $\widetilde{P}_\gamma$ in (179) is logrithmically singular (while for the Laplace equation, it is smooth). High-order discretization schemes for such integral equations can be found in the literature (see, for example [1,12,20]).

Needless to say, three-dimensional versions of most problems of mathematical physics are of more immediate applied interest than their two-dimensional versions. Thus, the results of this paper should be viewed as a model for the investigation of the Dirichlet problem for the Laplace equation (or some other elliptic PDE) in three dimensions, with the data specified on an open surface S. When the boundary S is smooth, the transition is fairly straightforward; it becomes more involved when S itself has corners. Both cases are presently under investigation.

Acknowledgements

The authors would like to thank the (anonymous) referees; all three made suggestions that the authors found useful and incorporated in the paper.

References

[1] B. Alpert, High-order quadratures for integral operators with singular kernels, J. Comput. Appl. Math. 60 (1995) 367–378.
[2] M. Abramowitz, I. Stegun (Eds.), Handbook of Mathematical Functions, Dover, New York, 1965.
[3] M.P. Do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, NJ, 1976.
[4] A. Friedman, Partial Differential Equations, Krieger, Huntington, 1976.
[5] D. Gottlieb, S.A. Orszag, Numerical Analysis of Spectral Methods: Theory and Applications, sixth ed., SIAM, Philadelphia, 1993.
[6] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series, and Products, fifth ed., Academic Press, New York, 1994.
[7] L. Greengard, V. Rokhlin, A new version of the fast multipole method for the Laplace equation in three dimensions, Acta Numer. 6 (1997) 229–269.
[8] J. Hadamard, Lectures on the Cauchy's Problem in Linear Partial Differential EquatiDover, Dover, New York, 1952.
[9] J. Helsing, On the numerical evalution of stress intensity factors for an interfaced crack of a general shape, Int. J. Numer. Meth. Eng. 44 (5) (1999) 729–741.
[10] J. Helsing, A. Jonsson, Stress calculations on multiply connected domains, J. Comput. Phys. 176 (2) (2002) 456–482.
[11] J. Helsing, G. Peters, Integral equation methods and numerical solutions of crack and inclusion problems in planar elastostatics, SIAM J. Appl. Math. 59 (3) (1999) 965–982.
[12] S. Kapur, V. Rokhlin, High-order corrected quadrature rule for singular functions, SIAM J. Numer. Anal. 34 (1997) 1331–1356.
[13] P. Kolm, V. Rokhlin, Numerical quadratures for singular and hypersingular integrals, Comput. Math. Appl. 41 (3–4) (2001) 327–352.
[14] P. Kolm, V. Rokhlin, Quadruple and Octuple Layer Potentials in Two Dimensions I: Analytical Apparatus, Tech. Rep. Yale YALEU/DCS/RR-1176, Computer Science Department, Yale University, 1999.
[15] R. Kress, Linear Integral Equations, second ed., Springer, New York, 1999.
[16] S.G. Mikhlin, Integral Equations and Their Applications to Certain Problems in Mechanics, Mathematical Physics and Technology, Pergamon Press, Oxford, 1957.
[17] N.I. Muskhelishvili, Singular Integral Equations, Dover, New York, 1991.
[18] H.L. Royden, Real Analysis, third ed., Prentice Hall, Englewood Cliffs, NJ, 1988.
[19] E.M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, 1971.
[20] J. Strain, Locally corrected multidimensional quadrature rules for singular functions, SIAM J. Sci. Comput. 16 (1995) 992–1017.
[21] F.G. Tricomi, Integral Equations, Dover, New York, 1957.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 195 (2004) 1–16
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Second kind integral equations for the classical potential theory on open surfaces II

Shidong Jiang *,[1], Vladimir Rokhlin

Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA

Received 11 July 2003; received in revised form 30 September 2003; accepted 1 October 2003

Abstract

A second kind integral equation formulation is presented for the Dirichlet problem for the Laplace equation in two dimensions, with the boundary conditions specified on a collection of open curves. The performance of the obtained apparatus is illustrated with several numerical examples. The formulation is a simplification of the equation previously constructed by the authors.

Keywords: Open surface problems; Laplace equation; Second kind integral equation; Dirichlet problem

AMS: 65R10; 77C05

1. Introduction

Integral equations have been one of principal tools for the numerical solution of scattering problems for more than 30 years, both in the Helmholtz and Maxwell environments. Historically, most of the equations used have been of the first kind, since numerical instabilities associated with such equations have not been critically important for the relatively small-scale problems that could be handled at the time.

The combination of improved hardware with the recent progress in the design of "fast" algorithms has changed the situation dramatically. Condition numbers of systems of linear algebraic equations resulting from the discretization of integral equations of potential theory have become critical, and the simplest way to limit such condition numbers is by starting with second kind integral equations. Hence, increasing interest in reducing scattering problems to systems of *second kind* integral equations on the boundaries of the scatterers.

* Corresponding author.
E-mail addresses: shidong@cs.yale.edu (S. Jiang), rokhlin@cs.yale.edu (V. Rokhlin).
[1] Supported in part by DARPA under Grant MDA972-00-1-0033, and by ONR under Grant N00014-01-1-0364.

0021-9991/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2003.10.001

During the last several years, satisfactory integral equation formulations have been constructed in both acoustic (Helmholtz equation) and electromagnetic (Maxwell's equations) environments, whenever all of the scattering surfaces are "closed" (i.e., scatterers have well-defined interiors, and have no infinitely thin parts). In this paper, we describe a second kind integral equation formulation for the Dirichlet problem for the Laplace equation with boundary data specified on a collection of "open" curves. We start with constructing the right inverse of the single layer potential operator on a line segment via simple analytic means; then we apply such operator as a preconditioner for the single layer potential operator on the curve considered to obtain a second kind integral operator.

Remark 1. In a recent paper [7], the authors construct a somewhat different procedure for the solution of problems of the classical potential theory with data specified on a collection of open surfaces. While the approach of the present paper is very similar to that of [7], in [7], the single layer potential is used to precondition the quadruple layer potential from the right; here, the quadruple layer potential is used to precondition the single layer potential from the right. For technical reasons, the latter leads to a drastically simplified numerical procedure (and also, requires simpler analysis); hence, this sequel to [7].

Remark 2. As observed by one of referees to this paper, a second kind integral equation is constructed in [11] (Chapter 16) in the Laplace environment. In [11], the solution of the Dirichlet problem is represented via the real part of the Cauchy's integral and the resulting boundary equation is a singular integral equation. A second kind integral equation is then obtained by applying the inverse operator of the Cauchy's integral operator from the left to both sides of the equation. However, the scheme of [11] cannot be easily extended either to three dimensions or to the Helmholtz equation in two dimensions, since it relies heavily on the harmonic property of the solution and the techniques of complex analysis.

Remark 3. As observed by another of referees to this paper, a closed surface enclosing very thin volumes presents difficulties closely related to those associated with open surfaces. This class of issues is not treated in this paper.

The layout of the paper is as follows. Section 2 contains an informal description of the procedure. In Section 3, the necessary mathematical and numerical preliminaries are introduced. In Sections 4, we present the principal analytic result of the paper. In Section 5, we describe a simple numerical implementation of the scheme. The performance of the algorithm is illustrated in Section 6 with several numerical examples. Finally, in Section 7 we discuss several generalizations of the approach.

2. Informal description of the procedure

In this section, we present an informal description of the procedure. We assume that $\gamma : [-1, 1] \to \mathbb{R}^2$ is a sufficiently smooth "open" (i.e., $\gamma(-1) \neq \gamma(1)$) curve with the parametrization

$$\gamma(t) = \tilde{\gamma}\left(\frac{L}{2} \cdot (t+1)\right), \tag{1}$$

where L is the total arc length of the curve, and $\tilde{\gamma} : [0, L] \to \mathbb{R}^2$ is the same curve parametrized by its arc length. The image of γ will be denoted by Γ. We consider the Dirichlet problem for the Laplace equation in two dimensions, with the boundary conditions specified on Γ, i.e.,

$$\begin{cases} \Delta u = 0 & \text{in } \mathbb{R}^2 \setminus \Gamma, \\ u = f & \text{on } \Gamma. \end{cases} \tag{2}$$

This problem has a unique bounded solution if the Dirichlet data f is sufficiently smooth (see, for example, [9, p. 121]). The purpose of this paper is to reduce the problem (2) to a second kind integral equation on Γ.

The tools of the classical potential theory by themselves do not lead to such an integral equation. Indeed, the standard prescription (see, for example, [9]) is to represent the solution of a Dirichlet problem by a *double* layer potential, and the solution of the Neumann problem by a *single* layer potential. In either case, the behavior of the singularity near the boundary is such that an integral equation of the second kind on Γ is obtained.

However, the classical procedure critically depends on Γ being a *closed* curve. Indeed, the potential of a double layer on the curve Γ experiences a jump when Γ is crossed; the magnitude of the jump is equal to the density of the double layer at the crossing point. This poses no problem when the curve is a closed one, since the potential is to be represented on only one (inner or outer) side of the curve. For an *open* curve, the potential has to be represented on both sides of the curve; and in most cases, the right-hand side f (viewed as the limiting value of the solution from both sides) has no jump across Γ. Thus, an attempt to represent the solution of (2) via a double layer potential results in a dipole density that is identically equal to zero.

One could attempt to represent the solution of (2) by a charge distribution on Γ. The resulting potential *is* continuous across Γ, and algorithms of this type have been constructed and used numerically (see, for example, [6]). However, the resulting integral equation is of the first kind (though, fortunately, with a logarithmically singular kernel), with all the usual numerical disadvantages. Another option is to use the quadruple layer potential of the form

$$R(\sigma)(x) = \int_{-1}^{1} \frac{\partial^2}{\partial N(t)^2} (\log ||x - \gamma(t)||) \cdot \sigma(t)\, dt, \tag{3}$$

with $N(t)$ the unit normal to Γ at $\gamma(t)$; the resulting equation is not an integral equation at all, containing a part that is actually a distribution. In engineering literature, such objects are known as "hypersingular integral equation". Satisfactory procedures have been constructed for their numerical solution (see, for example, [3,10,12]); however, these are not as simple or as stable as the many methods available for the solution of second kind integral equations.

This paper is based on the observation that when the curve is the line segment $I = [-1, 1]$, the right inverse of the single layer potential operator (denoted by S_I^{-1}) can be constructed by simple analytic means, where the single layer potential operator $S_I : L^1[-1, 1] \to C[-1, 1]$ is defined by the formula

$$S_I(\sigma)(x) = \int_{-1}^{1} \log |x - t| \cdot \sigma(t)\, dt. \tag{4}$$

Furthermore, if S_I^{-1} is used as a preconditioner for the single layer potential operator $S_\gamma : L^1[-1, 1] \to C(\mathbb{R}^2)$ on Γ defined by the formula

$$S_\gamma(\sigma)(z) = \int_{-1}^{1} \log |z - \gamma(t)| \cdot \sigma(t)\, dt, \tag{5}$$

i.e., the solution of the problem (2) is represented in the form

$$u(x) = S_\gamma \circ S_I^{-1}(\eta)(x), \tag{6}$$

then the resulting boundary integral equation is of the *second kind.*

Remark 4. A stable second kind integral equation formulation has also been developed for the problem (2) in [7]. Two key observations used in [7] are: first, the product of the quadruple layer potential operator and the single layer potential operator is a second kind integral operator for the case of a closed curve; second, the case of a line segment can be solved analytically. The integral representation for the solution of the problem (2) in [7] is of the form

$$u(x) = Q_\gamma \circ S_I \circ (Q_I \circ S_I)^{-1}(\eta)(x), \tag{7}$$

where Q_γ is the sum of a quadruple layer potential and a weighted double layer potential with the weight equal to the curvature, S_I is the single layer potential operator for the line segment $I = [-1, 1]$, and $(Q_I \circ S_I)^{-1}$ is (in the appropriate sense) the right inverse of $Q_I \circ S_I$. The approach of this paper differs from that of [7] in that the roles of Q and S are interchanged, leading to a simpler scheme. Indeed, straightforward analysis shows that the representation (6) is equivalent to

$$u(x) = S_\gamma \circ Q_I \circ (S_I \circ Q_I)^{-1}(\eta)(x). \tag{8}$$

In other words, the solution of (2) is represented by a single layer potential on Γ preconditioned by the quadruple layer potential for the line segment I, with a further preconditioning by the right inverse of $S_I \circ Q_I$ to eliminate the singularities at the end points.

3. Analytical preliminaries

In this section, we summarize several results from classical and numerical analysis to be used in the remainder of the paper. Detailed references are given in the text.

3.1. Chebyshev polynomials and Chebyshev approximation

Chebyshev polynomials are frequently encountered in numerical analysis. As is well known, Chebyshev polynomials of the first kind $T_n : [-1, 1] \to \mathbb{R}$ $(n \geqslant 0)$ are defined by the formula

$$T_n(x) = \cos(n \arccos(x)), \tag{9}$$

and are orthogonal with respect to the inner product

$$(f, g) = \int_{-1}^{1} f(x) \cdot g(x) \cdot \frac{1}{\sqrt{1 - x^2}}\, dx. \tag{10}$$

Chebyshev polynomials of the second kind $U_n : [-1, 1] \to \mathbb{R}$ $(n \geqslant 0)$ are defined by the formula

$$U_n(x) = \frac{\sin((n+1)\arccos(x))}{\sin(\arccos(x))}, \tag{11}$$

and are orthogonal with respect to the inner product

$$(f, g) = \int_{-1}^{1} f(x) \cdot g(x) \cdot \sqrt{1 - x^2}\, dx. \tag{12}$$

The Chebyshev nodes x_i of degree N are the zeros of T_N defined by the formula

$$x_i = \cos\frac{(2i+1)\pi}{2N}, \quad i = 0, 1, \ldots, N-1. \tag{13}$$

For a sufficiently smooth function $f : [-1, 1] \to \mathbb{R}$, its Chebyshev expansion is defined by the formula

$$f(x) = \sum_{k=0}^{\infty} C_k \cdot T_k(x), \tag{14}$$

with the coefficients C_k given by the formulae

$$C_0 = \frac{1}{\pi}\int_{-1}^{1} f(x)\cdot T_0(x)\cdot (1-x^2)^{-1/2}\,dx, \tag{15}$$

and

$$C_k = \frac{2}{\pi}\int_{-1}^{1} f(x)\cdot T_k(x)\cdot (1-x^2)^{-1/2}\,dx, \tag{16}$$

for all $k \geqslant 1$. We will also denote by P_f^N the order $N-1$ Chebyshev approximation to the function f on the interval $[-1,1]$, i.e., the (unique) polynomial of order $N-1$ such that $P_f^N(x_i) = f(x_i)$ for all $i = 0,1,\ldots,N-1$, with x_i the Chebyshev nodes defined by (13).

The following lemma provides an error estimate for the Chebyshev approximation (see, for example, [4]).

Lemma 5. *If $f \in C^k[-1,1]$ (i.e., f has k continuous derivatives on the interval $[-1,1]$), then for any $x \in [-1,1]$,*

$$\left|P_f^N(x) - f(x)\right| = O\left(\frac{1}{N^k}\right). \tag{17}$$

In particular, if f is infinitely differentiable, then the Chebyshev approximation converges superalgebraically (i.e., faster than any finite power of $1/N$ as $N \to \infty$).

3.2. Miscellaneous results

In this section, we collect several results from classical analysis to be used subsequently. Lemma 6 lists two standard definite integrals; both can be found (in a somewhat different form) in [5]. Lemma 7 states a standard fact from classical potential theory; it can be found in [9]. Finally, Lemma 8 states that if the curve γ is sufficiently smooth, then the restriction of the kernel of the operator $S_\gamma - S_I$ on Γ is also smooth (see (4), (5) for the definitions of S_I and S_γ).

Lemma 6. *For any $x \in (-1,1)$,*

$$\int_{-1}^{1} \log|x-t|\cdot\frac{1}{\sqrt{1-t^2}}\,dt = -\pi\cdot\log 2, \tag{18}$$

and

$$\text{p.v.}\int_{-1}^{1}\frac{1}{x-t}\cdot U_{n-1}(t)\cdot\sqrt{1-t^2}\,dt = \pi\cdot T_n(x), \tag{19}$$

for any $n \geqslant 1$.

Lemma 7. *Suppose that $\gamma : [-1,1] \to \mathbb{R}^2$ is a sufficiently smooth open regular curve with the parametrization (1), and that the function $\sigma \in L^1[-1,1]$ satisfies the condition*

$$\int_{-1}^{1}\sigma(t)\,dt = 0. \tag{20}$$

Then the function $u : \mathbb{R}^2 \to \mathbb{R}$ defined by the formula

$$u(x) = \int_{-1}^{1}\log|x-\gamma(t)|\cdot\sigma(t)\,dt \tag{21}$$

is bounded in $\mathbb{R}^2$.

Lemma 8. *Suppose that $\gamma \in C^{k+1}[0,L]$ $(k \geqslant 1)$ is an open regular curve parametrized by its arc length in $\mathbb{R}^2$. Suppose further that the function $r : [0,L] \times [0,L] \to \mathbb{R}$ is defined by the formula*

$$r(x,t) = \begin{cases} \log|\gamma(x) - \gamma(t)| - \log|x - t|, & x \neq t, \\ 0, & x = t. \end{cases} \tag{22}$$

Then $r \in C^k([0,L] \times [0,L])$.

Proof. Since γ is parametrized by its arc length, we have

$$|\gamma'(x)| = 1, \tag{23}$$

for all $x \in [0,L]$. Combining (22), (23), we observe that

$$r(x,t) = \log|h(x,t)|, \tag{24}$$

where the function $h : [0,L] \times [0,L] \to \mathbb{R}^2$ is defined by the formula

$$h(x,t) = \begin{cases} \frac{\gamma(x)-\gamma(t)}{x-t}, & x \neq t, \\ \gamma'(x), & x = t. \end{cases} \tag{25}$$

Obviously, h is k times continuously differentiable for $\gamma \in C^{k+1}[0,L]$ by Taylor's Theorem. Furthermore, since $\gamma(x) \neq \gamma(t)$ if $x \neq t$, and $|\gamma'(x)| = 1$ for all $x \in [0,L]$, we have

$$|h(x,t)| \neq 0 \quad \text{for all } (x,t) \in [0,L] \times [0,L]. \tag{26}$$

Therefore, the function $r = \log|h|$ is also k times continuously differentiable in $[0,L] \times [0,L]$. □

4. Analytical apparatus

4.1. Right inverse of the single layer potential operator on the line segment

The purpose of this section is Theorem 10, providing the right inverse of the single layer potential operator on the line segment $I = [-1,1]$. The construction is based on an elementary integral identity stated in Lemma 9.

Lemma 9. *For any $x \in (-1,1)$,*

$$\int_{-1}^{1} \log|x - t| \cdot \frac{T_0(t)}{\sqrt{1-t^2}}\,dt = -\pi \cdot \log 2 \cdot T_0(x), \tag{27}$$

and

$$\int_{-1}^{1} \log|x - t| \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\,dt = -\frac{\pi}{n} \cdot T_n(x), \tag{28}$$

for any $n \geqslant 1$.

Proof. (27) directly follows from the combination of the identity (18) and the fact that $T_0(x) = 1$ for all $x \in [-1,1]$. To prove (28), we integrate by parts once, obtaining

$$\int_{-1}^{1} \log|x-t| \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\,dt = -\frac{1}{n}\text{p.v.}\int_{-1}^{1} \frac{1}{x-t} \cdot U_{n-1}(t) \cdot \sqrt{1-t^2}\,dt. \tag{29}$$

Now, (28) follows from the combination of (29), (19). □

Theorem 10. *Suppose that the linear operator $\widetilde{S} : C[-1,1] \to L^1[-1,1]$ is defined by its action on the functions T_n $(n \geqslant 0)$ via the formula*

$$\widetilde{S}(T_n)(x) = \begin{cases} -\dfrac{1}{\pi \cdot \log 2} \cdot \dfrac{T_0(x)}{\sqrt{1-x^2}}, & n = 0, \\ -\dfrac{n}{\pi} \cdot \dfrac{T_n(x)}{\sqrt{1-x^2}}, & n > 0. \end{cases} \tag{30}$$

Suppose further that the operator $S_I : L^1[-1,1] \to C[-1,1]$ is defined by the formula

$$S_I(\sigma)(x) = \int_{-1}^{1} \log|x-t| \cdot \sigma(t)\,dt. \tag{31}$$

Then

$$S_I \circ \widetilde{S} = \mathcal{I}, \tag{32}$$

with $\mathcal{I}$ the identity operator. In other words, $\widetilde{S}$ is the right inverse of S_I on the space of continuous functions.

Proof. Since T_n $(n \geqslant 0)$ form a basis for the space $C[-1,1]$, and the operators S_I, $\widetilde{S}$ are linear, we only need to prove that the identity

$$S_I \circ \widetilde{S}(T_n)(x) = T_n(x) \tag{33}$$

holds for all $n \geqslant 0$. Substituting (30) into (31) we obtain

$$S_I \circ \widetilde{S}(T_n)(x) = \begin{cases} -\dfrac{1}{\pi \cdot \log 2} \cdot \displaystyle\int_{-1}^{1} \log|x-t| \cdot \frac{T_0(t)}{\sqrt{1-t^2}}\,dt, & n = 0, \\ -\dfrac{n}{\pi} \cdot \displaystyle\int_{-1}^{1} \log|x-t| \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\,dt, & n > 0. \end{cases} \tag{34}$$

Combining (33), (34), we observe that it suffices to prove the identity

$$\int_{-1}^{1} \log|x-t| \cdot \frac{T_n(t)}{\sqrt{1-t^2}}\,dt = \begin{cases} -\pi \cdot \log 2 \cdot T_0(x), & n = 0, \\ -\dfrac{\pi}{n} \cdot T_n(x), & n > 0, \end{cases} \tag{35}$$

which directly follows from Lemma 9. □

4.2. Second kind integral equation formulation

In this section, we reduce Problem (2) to an integral equation of the second kind on the curve Γ; the results are summarized in Theorem 12. We start with defining the operator $\widetilde{S}_\gamma : C[-1,1] \to C(\mathbb{R}^2)$ via the formula

$$\widetilde{S}_\gamma(\sigma)(z) = S_\gamma \circ \widetilde{S}(\sigma)(z), \tag{36}$$

with S_γ, $\widetilde{S}$ defined by (5), (30), respectively. Combining (36) with Theorem 10, we easily see that for arbitrary smooth $\sigma : [-1,1] \to \mathbb{R}$ and $\gamma(x) \in \Gamma$,

$$\widetilde{S}_\gamma(\sigma)(\gamma(x)) = S_I \circ \widetilde{S}(\sigma)(x) + (S_\gamma - S_I) \circ \widetilde{S}(\sigma)(\gamma(x)) = \sigma(x) + (S_\gamma - S_I) \circ \widetilde{S}(\sigma)(\gamma(x)), \tag{37}$$

and the following theorem shows that the operator $P_\gamma = (S_\gamma - S_I) \circ \widetilde{S}$ is compact.

Theorem 11. *Suppose that $\gamma : [-1,1] \to \mathbb{R}^2$ is a sufficiently smooth open regular curve with the parametrization* (1). *Suppose further that the operator $P_\gamma : C[-1,1] \to C[-1,1]$ is defined by the formula*

$$P_\gamma(\sigma)(x) = (S_\gamma - S_I)\circ\widetilde{S}(\sigma)(\gamma(x)) = \int_{-1}^{1} (\log|\gamma(x) - \gamma(t)| - \log|x - t|) \cdot \widetilde{S}(\sigma)(t)\,dt, \tag{38}$$

with S_γ, S_I, $\widetilde{S}$ defined by (5), (31), (30), *respectively. Then P_γ is compact.*

Proof. By Lemma 8, the function $\widetilde{r} : [-1,1] \times [-1,1] \to \mathbb{R}$ defined by the formula

$$\widetilde{r}(x,t) = \log|\gamma(x) - \gamma(t)| - \log|x - t| \tag{39}$$

is k times continuously differentiable for any $\gamma \in C^{k+1}[-1,1]$. Obviously, if $\widetilde{r}$ is expanded into a double Chebyshev series

$$\widetilde{r}(x,t) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} K_{mn} T_m(x) T_n(t), \tag{40}$$

then there exists a positive number C such that

$$|K_{mn}| < \frac{C}{m^k \cdot n^k} \tag{41}$$

for any $m > 0$, $n > 0$. Now, for any $N > 0$, we will define the operator $P_N : C[-1,1] \to C[-1,1]$ by the formula

$$P_N(\sigma)(x) = \int_{-1}^{1} \widetilde{r}_N(x,t) \cdot \widetilde{S}(\sigma)\,dt, \tag{42}$$

with the function $\widetilde{r}_N : [-1,1] \times [-1,1] \to \mathbb{R}$ defined by the formula

$$\widetilde{r}_N(x,t) = \sum_{m=0}^{N}\sum_{n=0}^{N} K_{mn} T_m(x) T_n(t). \tag{43}$$

Obviously, P_N is a compact operator since its range is of finite dimensionality. Furthermore, P_N converges to P_γ as $N \to \infty$ by (41). Hence, P_γ is also a compact operator.

We will represent the solution of Problem (2) via the formula

$$u(x) = \widetilde{S}_\gamma(\sigma)(x) + A = \int_{-1}^{1} \log|x - \gamma(t)| \cdot \widetilde{S}(\sigma)(t)\,dt + A, \tag{44}$$

where A is a real constant to be determined. Combining Lemma 7 and Theorem 11, we obtain the principal result of this paper. □

Theorem 12. *Suppose that $\gamma : [-1,1] \to \mathbb{R}^2$ is a sufficiently smooth open regular curve with the parametrization* (1), *and that the function $f : [-1,1] \to \mathbb{R}$ is continuously differentiable. Suppose further that the continuous function $\sigma : [-1,1] \to \mathbb{R}$ and the coefficient A satisfy the equations*

$$\sigma(x) + P_\gamma(\sigma)(x) = f(x) - A, \tag{45}$$

$$\int_{-1}^{1} \sigma(x) \cdot \frac{1}{\sqrt{1 - x^2}}\,dx = 0, \tag{46}$$

with P_γ defined in (38). *Then the function $u : \mathbb{R}^2 \to \mathbb{R}$ defined by* (44) *is bounded in $\mathbb{R}^2$ and is the solution of the problem*

$$\begin{cases} \Delta u = 0 & \text{in } \mathbb{R}^2 \setminus \Gamma, \\ u = f & \text{on } \Gamma. \end{cases} \tag{47}$$

Remark 13. Obviously, the purpose of the constant A in the above theorem is to ensure the boundedness of the solution u of (2). In certain physical situations, the potentials of interest are not bounded at infinity, but rather grow logarithmically. In such cases, the solution to (2) assumes the form

$$u(x) = \widetilde{S}_\gamma(\sigma)(x), \tag{48}$$

with σ satisfying the integral equation

$$\sigma(x) + P_\gamma(\sigma)(x) = f(x). \tag{49}$$

5. Numerical algorithm

In this section, we construct a rudimentary numerical algorithm for the solution of the Dirichlet problem (47) via the Eqs. (45) and (46). Since the construction of the matrix and the solver of the resulting linear system are direct, the algorithm requires $O(N^3)$ work and $O(N^2)$ storage, with N the number of nodes on the boundary. While standard acceleration techniques (such as the Fast Multipole Method, etc.) could be used to improve these estimates, no such acceleration was performed, since the purpose of this section (as well as the following one) is to demonstrate the stability of the integral formulation and the convergence rate of a very simple discretization scheme.

By Theorem 12, the equations to be solved are (45) and (46), where the unknowns are the function σ and the real number A. To solve (45) and (46) numerically, we discretize the boundary into N Chebyshev nodes and approximate the unknown density σ by a finite Chebyshev series of the first kind,

$$\sigma(t) \simeq \sum_{k=0}^{N-1} C_k \cdot T_k(t), \tag{50}$$

with the coefficients C_k $(k = 0, \ldots, N-1)$ to be determined. In order to discretize (45), we start with observing that by (29), the action of the operator $\widetilde{S}$ on the function σ is described via the formula

$$\widetilde{S}(\sigma)(x) = \frac{1}{\sqrt{1-x^2}} \sum_{k=0}^{N-1} B_k \cdot C_k \cdot T_k(x), \tag{51}$$

where the coefficients B_k $(k = 0, \ldots, N-1)$ are given by the formulae

$$\begin{cases} B_0 = -\dfrac{1}{\pi \log 2}, \\ B_k = -\dfrac{k}{\pi}, \quad 1 \leqslant k \leqslant N-1. \end{cases} \tag{52}$$

Next, we approximate the kernel $\widetilde{r}(x,t)$ (see (40)) of the operator $S_\gamma - S_I$ with an expression of the form

$$\widetilde{r}(x,t) \simeq \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} K_{ij} \cdot T_i(x) \cdot T_j(t). \tag{53}$$

Clearly, the coefficients K_{ij} have to be determined numerically, since the curve Γ is user-specified, and is unlikely to have a convenient analytical expression. Thus, we obtain the coefficients K_{ij} by first constructing the $N \times N$ matrix $R = (\tilde{r}(x_i, t_j))$ $(i, j = 0, 1, \ldots, N-1)$ with x_i, t_j the Chebyshev nodes defined by (13) then converting R into the matrix $K = (K_{ij})$ $(i, j = 0, 1, \ldots, N-1)$ by the formula

$$K = U \cdot R \cdot U^T, \tag{54}$$

with $N \times N$ matrix $U = (U_{ij})$ defined by the formula

$$\begin{cases} U_{0j} = \frac{1}{N} \cdot T_0(x_j), & j = 0, 1, \ldots, N-1, \\ U_{ij} = \frac{2}{N} \cdot T_i(x_j), & i = 1, \ldots, N-1, \quad j = 0, 1, \ldots, N-1, \end{cases} \tag{55}$$

Finally, we approximate the prescribed Dirichlet data f by its Chebyshev approximation of order $N-1$

$$f(t) \simeq \sum_{k=0}^{N-1} \hat{f}_k \cdot T_k(t), \tag{56}$$

where the coefficients $\hat{f}_k$ can be obtained by first evaluating f at Chebyshev nodes x_i, then applying to it the matrix U defined by (55), i.e.,

$$\hat{f}_k = \sum_{i=0}^{N-1} U_{ki} \cdot f(x_i). \tag{57}$$

Combining (51), (53), (56),we discretize (45) into the equation

$$\tilde{A} \cdot \begin{pmatrix} C_0 \\ C_1 \\ \vdots \\ C_{N-1} \end{pmatrix} + A \cdot \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \vdots \\ \hat{f}_{N-1} \end{pmatrix}, \tag{58}$$

with $N \times N$ matrix $\tilde{A}$ defined by the formula

$$\tilde{A} = I_N + K \cdot B, \tag{59}$$

with I_N the $N \times N$ identity matrix, and B the diagonal matrix defined by the formula

$$B_{ij} = B_i \cdot \delta_{ij}. \tag{60}$$

Furthermore, (46) leads to the equation

$$C_0 = 0, \tag{61}$$

Finally, (58) and (61) together form a linear system of dimension $N+1$ to be solved.

Remark 14. The generalization of the above scheme to the case of several disjoint open curves is straightforward, and has been implemented by the authors (see Example 4 in Section 6).

6. Numerical examples

A FORTRAN code has been written implementing the algorithm described in the preceding section. In this section, we demonstrate the performance of the scheme with several numerical examples. We consider

the problem in electrostatics: the boundary is made of conductor and grounded, the electric field incident on the boundary is generated by the sources outside the boundary. For these examples, we plot the equipotential lines of the total field and present tables showing the convergence rate of the algorithm.

Remark 15. In the examples below, the problems to be solved via the procedure of the preceding section have no simple analytical solution. Thus, we tested the accuracy of our procedure by evaluating our solution via the formula (44) at a large number M of nodes on the boundary Γ (in our experiments, we always used $M = 2000$), and comparing it with the analytically evaluated right-hand side. We did not need to verify the fact that our solutions satisfy the Laplace equation, since this follows directly from the representation (44).

In each of those tables, the first column contains the total number N of nodes in the discretization of each curve. The second column contains the condition number of the linear system. The third column contains the relative L^2 error of the numerical solution as compared with the analytically evaluated Dirichlet data on the boundary. The fourth column contains the maximum absolute error on the boundary. In the last two columns, we list the errors of the numerical solution as compared with the numerical solution with twice the number of nodes, where the solution is evaluated at 1000 equispaced points on a circle of radius 3.3 centered at the origin; the fifth column contains the relative L^2 error, and the sixth column contains the maximum absolute error.

Example 1. In this example, the boundary is the line segment parametrized by the formula

$$\begin{cases} x(t) = t, \\ y(t) = -0.2, \end{cases} \quad -1 \leqslant t \leqslant 1. \tag{62}$$

The Dirichlet data are generated by a unit charge at $(0,0)$. The numerical results are shown in Table 1. The source, curve and equipotential lines are plotted in Fig. 1.

Example 2. In this example, the boundary is a sinusoidal arc parametrized by the formula

$$\begin{cases} x(t) = 0.5t, \\ y(t) = \cos(t), \end{cases} \quad -\frac{3\pi}{2} \leqslant t \leqslant \frac{3\pi}{2}. \tag{63}$$

The Dirichlet data are generated by one positive charge of unit strength at (0, 1.5) and another negative charge of unit strength at (0, 0). The numerical results are shown in Table 2. The sources, curve and equipotential lines are plotted in Fig. 2.

Table 1
Numerical results for Example 1

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
8	0.200E+01	0.703E−01	0.178E+00	0.296E−02	0.528E−02
16	0.222E+01	0.759E−02	0.212E−01	0.641E−04	0.114E−03
32	0.212E+01	0.165E−03	0.486E−03	0.556E−07	0.991E−07
64	0.206E+01	0.147E−06	0.446E−06	0.835E−13	0.150E−12
128	0.203E+01	0.225E−12	0.690E−12	0.355E−15	0.222E−14
256	0.202E+01	0.935E−15	0.214E−13	0.343E−15	0.200E−14

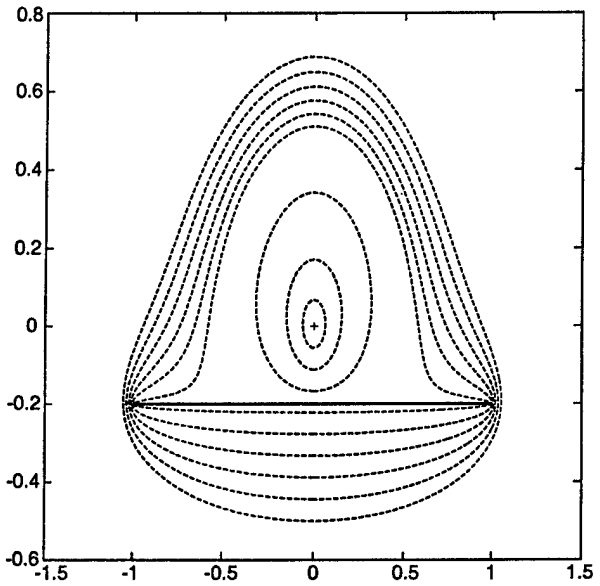


Fig. 1. Source, curve, and equipotential lines for Example 1.

Table 2
Numerical results for Example 2

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
32	0.195E+01	0.271E−01	0.864E−01	0.658E−02	0.469E−02
64	0.187E+01	0.240E−02	0.847E−02	0.146E−03	0.104E−03
128	0.182E+01	0.422E−04	0.157E−03	0.135E−06	0.955E−07
256	0.179E+01	0.307E−07	0.117E−06	0.245E−12	0.173E−12
512	0.178E+01	0.431E−13	0.160E−12	0.971E−15	0.133E−14
1024	0.177E+01	0.304E−14	0.450E−13	0.941E−15	0.122E−14

Example 3. In this example, the boundary is a spiral parametrized by the formula

$$\begin{cases} x(t) = t\cos(3.3\pi t) - 0.1, \\ y(t) = t\sin(3.3\pi t), \end{cases} \quad 0.2 \leqslant t \leqslant 3.2. \tag{64}$$

The Dirichlet data are generated by a unit charge at $(0,0)$. The numerical results are shown in Table 3. The source, curve and equipotential lines are plotted in Fig. 3.

Example 4. In this example, we consider the case of several open curves. The boundary consists of three elliptic arcs parametrized by the formulae

$$\begin{cases} x_1(t) = -t\cos(3.3\pi t) - 1.45, \\ y_1(t) = -t\sin(3.3\pi t) + 0.55, \end{cases} \quad 0.2 \leqslant t \leqslant 1.2, \tag{65}$$

$$\begin{cases} x_2(t) = t\cos(3.3\pi t) - 0.1, \\ y_2(t) = t\sin(3.3\pi t) - 1.2, \end{cases} \quad 0.2 \leqslant t \leqslant 1.2, \tag{66}$$

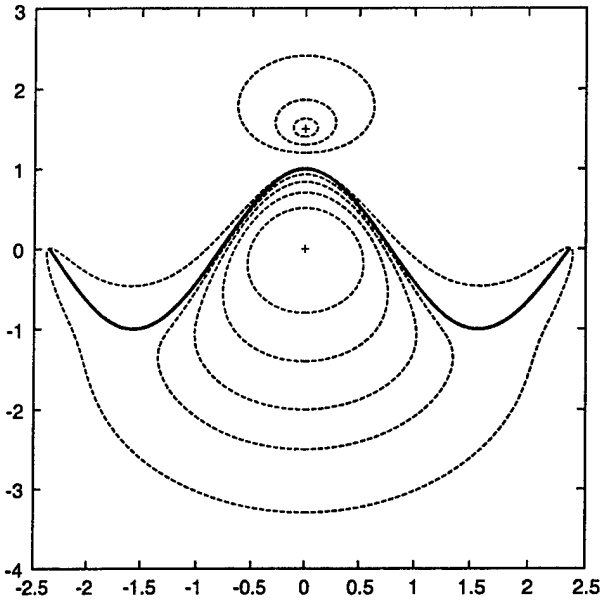


Fig. 2. Sources, curve, and equipotential lines for Example 2.

Table 3
Numerical results for Example 3

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
32	0.704E+03	0.594E−01	0.125E+00	0.233E+00	0.685E−01
64	0.657E+02	0.108E−02	0.665E−02	0.417E−02	0.201E−02
128	0.523E+02	0.904E−04	0.653E−03	0.101E−03	0.575E−04
256	0.394E+02	0.213E−05	0.183E−04	0.179E−06	0.125E−06
512	0.279E+02	0.313E−08	0.272E−07	0.156E−11	0.123E−11
1024	0.196E+02	0.184E−13	0.147E−12	0.211E−13	0.933E−14

$$\begin{cases} x_3(t) = t\cos(3.3\pi t) + 1.25, \\ y_3(t) = t\sin(3.3\pi t) + 0.85, \end{cases} \quad 0.2 \leqslant t \leqslant 1.2. \tag{67}$$

The Dirichlet data are generated by four unit charges located at $(0,0)$, $(1.35,0.75)$, $(-1.55,0.75)$, $(0,-1.2)$. The numerical results are shown in Table 4, where N is the number of nodes on each curve. The sources, curves and equipotential lines are plotted in Fig. 4.

Remark 16. The above examples illustrate the superalgebraic convergence of the scheme for smooth data and curves (see Lemmas 5, 8). The number of nodes needed depends on the complexity of the underlying geometry and the smoothness of the prescribed data. The condition number of the resulting linear system is usually very low.

7. Conclusions and generalizations

We have presented a second kind integral equation formulation for the Dirichlet problem for the Laplace equation in two dimensions, with the boundary condition specified on a curve (consisting of one or

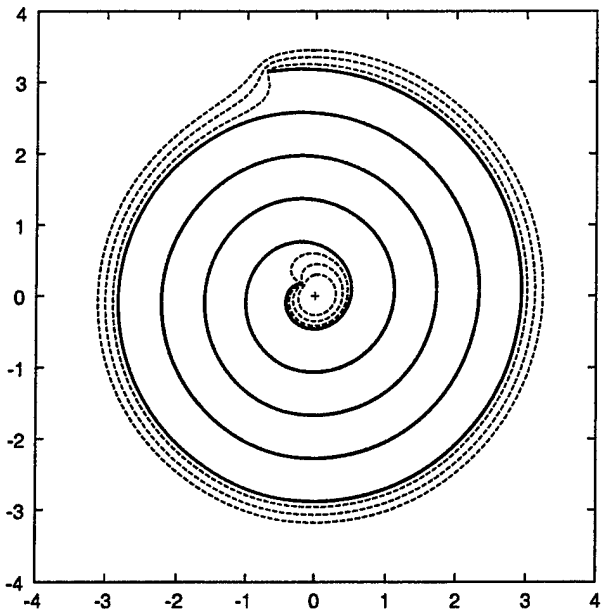


Fig. 3. Source, curve, and equipotential lines for Example 3.

Table 4
Numerical results for Example 4

N	K	$E^2(\Gamma)$	$E^\infty(\Gamma)$	$E^2(u)$	$E^\infty(u)$
8	0.204E+02	0.825E−01	0.370E+00	0.848E+01	0.451E−01
16	0.183E+02	0.180E−01	0.121E+00	0.259E+00	0.121E−02
32	0.145E+02	0.183E−02	0.131E−01	0.665E−03	0.355E−05
64	0.116E+02	0.355E−04	0.455E−03	0.738E−07	0.252E−09
128	0.963E+01	0.314E−07	0.353E−06	0.302E−11	0.232E−13
256	0.851E+01	0.511E−13	0.520E−12	0.269E−11	0.192E−13

more separate segments). The resulting numerical algorithm converges superalgebraically whenever both the boundary data and the curves are smooth.

In order to concentrate on the derivation and analysis of the integral formulation, we have chosen to use a very simple numerical scheme (see Section 5 above); the CPU time requirements of the procedure of Section 5 scale as n^3, with n the number of nodes in the discretization of the curve where the boundary condition is specified. A straightforward combination of the Fast Multipole Method (FMM), Fast Fourier Transform (FFT), and one of many standard iterative solvers yields an order $n \cdot \log(n)$ algorithm; such a scheme has been implemented, and will be reported at a later date. It is also possible to construct an order n scheme via the use of the FMM alone; according to the authors' estimates, for problems of practical size, this would offer no advantages over an FFT-based procedure.

Remark 17. In the iterative scheme outlined above, *each step* requires order $n \cdot \log(n)$ operations. Obviously, the complexity of the scheme also depends on the number of iterations needed to reach a required tolerance, which is to a large extent (though not entirely) determined by the spectral behavior of the discretized system. As observed by one of referees to this paper, a critical question for large-scale problems is

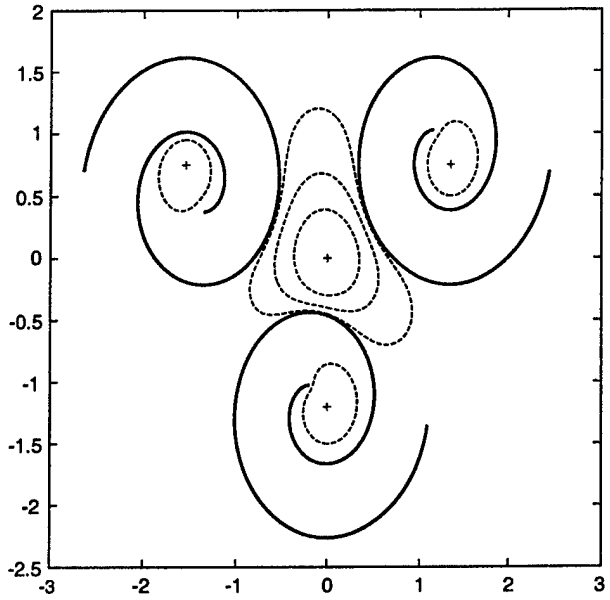


Fig. 4. Sources, curves, and equipotential lines for Example 4.

how the spectrum of the matrix A of the discretized system (or the spectrum of A^*A depending on the algorithm used) behaves as more and more reasonably separated curves of similar shapes are added to the geometry. This is currently under investigation.

The scheme of this paper can be applied almost without modification to elliptic PDEs other than the Laplace equation (such as Helmholtz equation, Yukawa equation, etc.). Indeed, the Green's function for any such equation has the form

$$G(x,y) = \phi(x,y) \cdot \log(||x-y||) + \psi(x,y), \tag{68}$$

with ϕ, ψ a pair of smooth functions (see, for example, [2]). When the procedure of Section 4 of this paper is applied to Green's function of the form (68), the result is unchanged, except for the change in the compact operator P_γ in (38). However, the convergence rate of the numerical scheme of Section 5 deteriorates drastically, since in this case the kernel K of the operator P_γ in (38) is logarithmically singular (while for the Laplace equation, it is smooth). High-order discretization schemes for such integral equations can be found in the literature (see, for example, [1,8,13]).

Finally, most results of this paper admit generalizations to two-dimensional surfaces in $\mathbb{R}^3$; while the necessary analytical apparatus is more involved, the results are very similar to those obtained here. Specifically, the product of a hypersingular integral operator on an open surface in $\mathbb{R}^3$ with the single layer potential operators (either from the left or from the right) is an integral operator of the second kind, except for simple corrections at the boundary of the surface. Such a scheme in three dimensions is being implemented, and will be reported at a later date.

Acknowledgements

The authors thank the referees of this paper for their helpful suggestions.

References

[1] B. Alpert, High-order quadratures for integral operators with singular kernels, J. Comput. Appl. Math. 60 (1995) 367–378.
[2] A. Friedman, Partial Differential Equations, Krieger, Huntington, 1976.
[3] J. Giroire, J.C. Nedelec, Numerical solution of an exterior Neumann problem using a double layer potential, Math. Comp. 32 (144) (1978) 973–990.
[4] D. Gottlieb, S.A. Orszag, Numerical Analysis of Spectral Methods: Theory and Applications, sixth ed., SIAM, Philadelphia, PA, 1993.
[5] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series, and Products, fifth ed., Academic Press, New York, 1994.
[6] G.C. Hsiao, R.C. MacCamy, Solution to boundary value problems by integral equations of the first kind, SIAM Rev. 15 (1973) 687–705.
[7] S. Jiang, V. Rokhlin, Second Kind Integral Equations for Scattering by Open Surfaces I: Analytical Apparatus, Tech. Rep. Yale YALEU/DCS/TR-1233, Computer Science Department, Yale University, 2002.
[8] S. Kapur, V. Rokhlin, High-order corrected quadrature rule for singular functions, SIAM J. Numer. Anal. 34 (1997) 1331–1356.
[9] R. Kress, Linear Integral Equations, second ed., Springer, Berlin, 1999.
[10] M.N. Leroux, Equations Intégrales pour le probleme du potentiel életrique dans le plan, C. R. Acad. Sci. Paris Ser. Math. A 178 (1974) 541–544.
[11] N.I. Muskhelishvili, Singular Integral Equations, second ed., Dover, New York, 1992.
[12] J.C. Nedelec, Curved finite element methods for the solution of singular integral equations on surface in R^3, Comput. Meth. Appl. Mech. Eng. 8 (1974) 61–80.
[13] J. Strain, Locally corrected multidimensional quadrature rules for singular functions, SIAM J. Sci. Comput. 16 (1995) 992–1017.